

UNIVERSIDADE FEDERAL DO PARANÁ

ANA BEATRIZ SCHIKOWSKI

**ESTIMATIVA DO VOLUME E DA FORMA DO FUSTE UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA**

CURITIBA

2016

ANA BEATRIZ SCHIKOWSKI

**ESTIMATIVA DO VOLUME E DA FORMA DO FUSTE UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Florestal, no Curso de Pós-Graduação em Engenharia Florestal, Setor de Ciências Agrárias, Universidade Federal do Paraná.

Orientador: Profa. Dra. Ana Paula Dalla Corte
Coorientador: Prof. Dr. Carlos Roberto Sanquetta

CURITIBA

2016

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

Schikowski, Ana Beatriz

Estimativa do volume e da forma do fuste utilizando técnicas de aprendizagem de máquina / Ana Beatriz Schikowski. – Curitiba, 2016.
71 f. : il.

Orientador: Prof.^a Dr.^a Ana Paula Dalla Corte

Coorientador: Prof. Dr. Carlos Roberto Sanquetta

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 11/03/2016.

Área de concentração: Manejo Florestal.

1. Troncos (Botânica) - Medição. 2. Árvores - Medição. 3. Manejo florestal. 4. Mineração de dados (Computação). 5. Inteligência artificial. 6. Teses. I. Dalla Corte, Ana Paula. II. Sanquetta, Carlos Roberto. III. Universidade Federal do Paraná, Setor de Ciências Agrárias. IV. Título.

CDD – 634.9
CDU – 634.0.521

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ANA BEATRIZ SCHIKOWSKI**, intitulada: **"ESTIMATIVA DO VOLUME E DA FORMA DO FUSTE UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA"**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua Aprovação.

Curitiba, 11 de Março de 2016.


Prof ANA PAULA DALLA CORTE (UFPR)
(Presidente da Banca Examinadora)


Prof JAIME WOJCIECHOWSKI (UFPR)


Prof JULIO EDUARDO ARCE (UFPR)



À minha família, dedico.

AGRADECIMENTOS

Agradeço à minha família, especialmente minha mãe Ruth pela dedicação e imensurável amor. Ao meu filho Vítor, por me mostrar o quão incrível é a aventura de ser mãe, assim como desafiador. À minha irmã Carol pelo incondicional apoio e puxões de orelhas certos, ao meu cunhado Eduardo e a minha sobrinha Helena. Sem dúvida a família é a razão de todo o esforço empreendido.

Agradeço ao meu mestre Doutor Daisaku Ikeda e aos companheiros budistas, da Associação Brasil Soka Gakkai Internacional, pelos incentivos recebidos ao longo dessa jornada.

Aos meus orientadores, Ana Paula Dalla Corte e Carlos Roberto Sanquetta, que ao longo dos cinco anos que trabalhei no laboratório junto a eles, pude sem dúvida, crescer técnica, profissional e pessoalmente.

Aos colegas de BIOFIX pelo companheirismo, cafés e artigos em conjunto. Em especial as amigas Lara Melo, Ângela Hentz, Tauane Garcia e Bruna Nascimento pelo apoio e amizade. Ao Alexandre Behling, que além de ser uma pessoa de excepcional inteligência, sempre foi benevolente o suficiente para dividir seu conhecimento a qualquer momento.

Aos amigos do Laboratório de Incêndios Florestais pelos cafezinhos e conversas sobre os mais diversos assuntos, procrastinações que muitas vezes melhoraram o rendimento dos artigos no turno da tarde. Especialmente a Bruna Kovalsky e Tatiane Ho, companheiras desde a graduação. Ao Igor Takashina que além do companheirismo e amizade, sempre cuidou para que meu computador estivesse sempre em perfeito funcionamento!

Aos grandes mestres da graduação e da pós-graduação pela preciosa contribuição na minha formação.

Aos demais colegas de graduação e pós pela troca de experiência ao decorrer desses oito anos de estudo.

Por fim, quero agradecer a CAPES pela concessão da bolsa de estudos e também a sociedade brasileira pelo financiamento, sem estes não seria possível o desenvolvimento deste trabalho.

“A educação é um veículo para desenvolver no próprio caráter o espírito de abraçar e engrandecer a vida dos outros. Dessa maneira, a missão orgulhosa daqueles que puderam receber educação deve ser servir, de alguma forma, àqueles que não tiveram essa oportunidade.”

Daisaku Ikeda

RESUMO

A acurácia e precisão na avaliação quantitativa de povoamentos para fins comerciais são fundamentais, especialmente a estimativa do volume individual. Nesse contexto, as equações de volume e funções de afilamento são importantes ferramentas para a estimativa do volume individual, ambos com métodos tradicionalmente aplicados e sólida teoria. Por outro lado, a inovação matemática é muito dinâmica, com diversos outros recursos disponíveis para o estudo de aproximação de função (regressão), métodos esses que ainda carecem de estudos de base, podendo potencialmente melhorar as estimativas no campo de modelagem florestal. Assim, o trabalho tem como objetivo principal analisar a acurácia de técnicas de aprendizado de máquina em relação a um modelo volumétrico e a uma função de afilamento, para a espécie *Acacia mearnsii* De Wild. Foram utilizados dados de cubagem *Acacia mearnsii*, variando de 1 a 10 anos. A base de dados foi dividida em 60% para ajuste e o restante para validação. Foram ajustadas equações com o modelo volumétrico de Schumacher e Hall e com a função de afilamento de Hradetzky, comparados com três algoritmos de aprendizado de máquina: *k*-vizinho mais próximo (*k*-NN), *Random Forest* (RF) e Redes Neurais Artificiais (RNA) para a estimativa do volume total e diâmetro referente à altura relativa. Os modelos foram ranqueados conforme estatísticas de erros, bem como observadas as distribuições destes. Para a estimativa do volume em função do dap e altura, a RNA e o modelo de Schumacher e Hall apresentaram melhores resultados no ranqueamento do que o *k*-NN e RF. Os métodos de aprendizado de máquina aplicados se mostraram mais acurados que o polinômio de Hradetzky para estimativas da forma da árvore, tais como o diâmetro ao longo do fuste e volume total. Os modelos de AM se mostraram adequados como alternativa na modelagem tradicionalmente aplicada na mensuração florestal, contudo a sua utilização deve ser cuidadosa devida a maior possibilidade de supertreinamento a base de ajuste.

Palavras-chave: Modelagem. Inteligência Artificial. Mineração de dados.

ABSTRACT

Accuracy and precision are essential topics when it comes to the quantitative evaluation of a forest stand with commercial purposes, especially in regards to the estimation of individual volume. In one hand, due to its solid theory, volume equations and taper functions are important tools for estimating individual volume. On the other hand, the mathematical breakthrough is dynamic, having several resources for the study of approximation functions (regression). However, these new methods still lack baseline studies and may potentially improve estimates in the forest modeling field. Under this circumstances, this study aims to analyze the accuracy of machine learning techniques in regards to a volumetric model and a taper function for the species *Acacia mearnsii* De Wild. *Acacia mearnsii* scaling data, from 1 to 10 years, were used in order to achieve the objective of this work. The database was divided into 60% for adjustment and the remainder for validation. Equations were adjusted with the volume model of Schumacher and Hall and the taper function Hradetzky and were compared to three machine learning algorithms: nearest neighbor models (k-NN), Random Forest (RF) and Artificial Neural Networks (ANN) to estimate of the total volume and diameter concerning relative height. The models were ranked according to its statistics errors and distributions. ANN and Schumacher and Hall model showed better results in regards to the estimation of volume as a function of dbh and height. The applied machine learning methods were more accurate than Hradetzky polynomial to estimates of tree shape, such as the diameter along the stem and total volume. The machine learning models were satisfactory as an alternative to the traditional methods in forest measurement. However, due to its higher possibility of overtraining the adjustment basis, it should be carefully used.

Keywords: Modelling. Artificial Intelligence. Data mining.

LISTA DE FIGURAS

FIGURA 1 – IMPACTO DO VALOR DE K NO ALGORITMO K-NN	20
FIGURA 2 – REPRESENTAÇÃO DE UMA ÁRVORE DE REGRESSÃO	23
FIGURA 3 – MODELO SIMPLIFICADO DE UM NEURÔNIO E SUA ESTRUTURA.	26
FIGURA 4 – REPRESENTAÇÃO GRÁFICA DA ARQUITETURA DE UMA RNA	27
FIGURA 5 – LOCALIZAÇÃO ESPACIAL DOS MUNICÍPIOS DOS PLANTIOS ESTUDADOS.....	30
FIGURA 7 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO VOLUME TOTAL EM FUNÇÃO DO DAP E ALTURA TOTAL.....	43
FIGURA 8 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO DIÂMETRO AO LONGO DO FUSTE.....	48
FIGURA 9 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO VOLUME TOTAL DO FUSTE	51

LISTA DE QUADROS

QUADRO 1 - COMPARAÇÃO DAS QUATRO TÉCNICAS DE MODELAGEM APLICADAS	54
---	----

LISTA DE TABELAS

TABELA 1 – DISTRIBUIÇÃO DA FREQUÊNCIA DOS INDIVÍDUOS AMOSTRADOS POR CLASSES IDADE E DE DIÂMETRO EM FUNÇÃO DOS GRUPOS DE AJUSTE E VALIDAÇÃO	31
TABELA 2 – COEFICIENTES E INDICADORES ESTATÍSTICOS DOS MODELOS VOLUMÉTRICOS AJUSTADOS POR CLASSE DE DAP	40
TABELA 3 – AVALIAÇÃO DAS CONDICIONANTES DA REGRESSÃO PARA OS RESÍDUOS DOS MODELOS DE SCHUMACHER E HALL AJUSTADOS POR CLASSE DE DAP.....	41
TABELA 4 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO VOLUME TOTAL DO FUSTE.....	42
TABELA 5 – COEFICIENTES, POTENCIAS SELECIONADAS PELO STEPWISE E INDICADORES ESTATÍSTICOS DAS FUNÇÕES DE AFILAMENTO AJUSTADAS POR CLASSE DE DAP	45
TABELA 6 – AVALIAÇÃO DAS CONDICIONANTES DA REGRESSÃO PARA OS RESÍDUOS DAS FUNÇÕES DE AFILAMENTO POR CLASSE DE DAP	46
TABELA 7 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO DIÂMETRO AO LONGO DO FUSTE	47
TABELA 8 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO VOLUME TOTAL DO FUSTE.....	50

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
2.1	OBJETIVO GERAL.....	14
2.2	OBJETIVOS ESPECÍFICOS	14
3	REVISÃO DE LITERATURA.....	15
3.1	VARIAÇÃO DA FORMA DO FUSTE	15
3.2	MODELAGEM DO VOLUME TOTAL	16
3.3	MODELAGEM DA FORMA DO FUSTE	16
3.4	INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA.....	18
3.4.1	<i>k</i> -Vizinho mais próximo	19
3.4.2	<i>Random Forest</i>	22
3.4.3	Redes Neurais Artificiais	25
4	MATERIAL E MÉTODOS.....	30
4.1	CARACTERIZAÇÃO DA ÁREA DE ESTUDO	30
4.2	CUBAGEM	31
4.3	EQUAÇÃO DE VOLUME	32
4.4	FUNÇÃO DE AFILAMENTO.....	32
4.5	ALGORITMOS DE APRENDIZADO DE MÁQUINA	34
4.5.1	<i>k</i> -Vizinho mais próximo	35
4.5.2	<i>Random Forest</i>	36
4.5.3	Redes neurais artificiais	36
4.6	AValiação dos Modelos.....	37
5	RESULTADOS E DISCUSSÃO	40
5.1	IA X FUNÇÃO DE VOLUME.....	40
5.1.1	Ajuste e teste das condicionantes de regressão do modelo de Schumacher e Hall.....	40

5.1.2	Estimativa do volume total obtido em função do dap e altura total	41
5.2	IA X FUNÇÃO DE AFILAMENTO.....	45
5.2.1	Ajuste e teste das condicionantes de regressão da função de afilamento polinômio de Hradetzky	45
5.2.2	Estimativa do diâmetro ao longo do fuste	46
5.2.3	Estimativa do volume total	50
5.3	ABORDAGEM COMPARATIVA DOS MÉTODOS ANALISADOS.....	53
6	CONCLUSÕES	57
7	RECOMENDAÇÕES PARA TRABALHOS FUTUROS	58
	REFERÊNCIAS.....	59

1 INTRODUÇÃO

A acurácia e precisão na avaliação quantitativa de povoamentos para fins comerciais são fundamentais, especialmente a mensuração do volume individual, sendo a cubagem o meio usual para esse fim. Como a quantificação por meio da cubagem é onerosa, se faz necessário aliar métodos indiretos para sua estimativa, sendo comumente utilizadas as equações de volume, funções de afilamento e fator de forma (SCHRÖDER et al., 2013), tendo como base as variáveis de fácil mensuração, como diâmetro a 1,30 m do solo (dap) e altura total.

Nesse contexto, as equações de volume e funções de afilamento são importantes ferramentas para a estimativa do volume individual. A primeira pela facilidade do ajuste e obtenção do volume com base em variáveis de fácil mensuração. A segunda por tratar-se de uma descrição matemática do perfil longitudinal de um tronco, sendo possível a obtenção de seu volume por meio da integração desta função. Destaque as funções de afilamento por permitirem a estimativa do diâmetro a uma altura qualquer do fuste, da altura comercial para um dado diâmetro, ou, ainda, do volume até um diâmetro mínimo qualquer (YOSHITANI JR et al., 2012), apresentando assim a vantagem de estimar o volume da madeira por sortimento, que se refere às diversas classes de uso da madeira.

Por outro lado, apesar de modelos consagrados na literatura e prática de empresas florestais, as equações de volume e funções de afilamento são modelos rígidos, sem flexibilidade e comportamento previamente definido. Nesse sentido, técnicas matemáticas alternativas como redes neurais artificiais vêm ganhando notoriedade no campo da estimação, dada a facilidade de aplicação e plasticidade no trato dos dados (BINOTI et al., 2014b; LEDUC et al., 2015), sendo potencialmente úteis na ciência florestal como uma alternativa à análise de regressão convencionalmente aplicada, dada à possibilidade de maior flexibilidade e modelagem de relações complexas entre variáveis.

Levando-se em consideração que tais métodos ainda carecem de estudos e fundamentações de base para a área florestal, se faz necessária a análise destes, para quem sabe assim melhorar a eficiência no campo de modelagem florestal. Essa melhora incide diretamente também no aumento da confiabilidade dos estimadores populacionais, como volume total por hectare e sortimento por hectare, acarretando

em maior segurança na tomada de decisões acerca de grandes empreendimentos florestais.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Este trabalho tem como objetivo principal analisar a acurácia de técnicas de aprendizado de máquina em relação a um modelo volumétrico e a uma função de afilamento, ambos tradicionalmente aplicados em mensuração e modelagem florestal, para a espécie *Acacia mearnsii* De Wild.

2.2 OBJETIVOS ESPECÍFICOS

Avaliar a acurácia das técnicas de aprendizado de máquina: vizinho mais próximo, *Random forest* e Redes Neurais Artificiais, para a estimativa do volume total em função do diâmetro e altura total para a espécie *Acacia mearnsii* De Wild.

Avaliar a acuracidade das técnicas de aprendizado de máquina: vizinho mais próximo, *Random forest* e Redes Neurais Artificiais, para a estimativa de diâmetro ao longo do fuste e volume total para a espécie *Acacia mearnsii* De Wild.

Comparar o desempenho, vantagens e desvantagens, das técnicas citadas com o modelo volumétrico de Schumacher e Hall (1933) e com a função de afilamento do polinômio de potências fracionárias e inteiras de Hradetzky (1976) para a espécie em questão, comumente aplicados nos trabalhos de inventários florestais.

Comparar o desempenho estatístico das técnicas de aprendizado de máquina citadas e dos modelos tradicionais (Schumacher e Hall e Hradetzky) com a cubagem real obtida pelo inventário florestal para a espécie *Acacia mearnsii* De Wild.

3 REVISÃO DE LITERATURA

3.1 VARIAÇÃO DA FORMA DO FUSTE

Historicamente a forma do fuste têm sido foco de diversos estudos e teorias na área de mensuração florestal, isso devida a sua alta relação com o volume individual da árvore. Estes estudos, por sua vez, são unânimes que não há uma forma padrão, mas sim que para cada espécie, em seus diversos estágios de desenvolvimento e condições ambientais, esta pode apresentar forma distinta. A forma do fuste pode ser definida como a taxa relativa de mudança do diâmetro do tronco com o aumento da altura da árvore (LARSON, 1963). Husch et al. (1982) complementam que a forma do fuste pode ser associada ao termo afilamento, que representa a diminuição do diâmetro com o aumento da altura.

A forma do tronco das árvores não pode ser considerada apenas com uma forma de um sólido específico, mas sim, como vários sólidos, a depender da porção considerada (HUSCH et al., 1982; FRIEDL, 1989). Segundo Husch et al. (1982), os troncos das árvores assumem formas de sólidos do tipo neilóide na base, parabolóide na porção intermediária e cone na extremidade final, sendo que ao se considerar o fuste como um todo, este não assumirá plenamente todos os sólidos citados, bem como, os pontos limítrofes desses sólidos são de difícil determinação, sendo, portanto de pouca importância prática. Por outro lado, desempenham grande valor teórico na definição de expressões matemáticas visando descrever a forma do fuste (FIGUEIREDO FILHO, 1991).

A forma do fuste varia de uma espécie para outra e também entre indivíduos da mesma espécie. Dentre os diversos fatores que influenciam a forma, destacam-se: tamanho da copa, hereditariedade, espaçamento, tratos culturais, posição sociológica, idade e sítio (LARSON, 1963; FINGER, 1992; PRODAN et al. 1997; MACHADO e FIGUEIREDO FILHO, 2006).

Devido a essa grande variação de forma inerente ao fuste, a determinação do seu volume se torna uma tarefa de difícil execução, necessitando, portanto, de métodos precisos e que sejam capazes de abranger as diversas formas que uma mesma espécie possa apresentar com a precisão requerida. De maneira geral, a forma do fuste das árvores vem sendo avaliada usando fatores de forma, quocientes de forma e funções de afilamento (KOHLE, 2013).

3.2 MODELAGEM DO VOLUME TOTAL

Segundo Figueiredo Filho et al. (1995), a modelagem do volume total pode ser realizada com certa facilidade e acuracidade empregando as equações de volume, geralmente ajustadas com base nas variáveis de fácil mensuração, como d_{ap} e altura total.

Assim, diversos modelos matemáticos foram elaborados para esse fim, porém, como Campos e Leite (2009) afirmam, o modelo de Schumacher e Hall (1933) tem seu uso consagrado devido às suas propriedades estatísticas, facilidade de ajuste e grande adaptabilidade as mais diversas bases de dados, gerando estimativas de boa qualidade.

3.3 MODELAGEM DA FORMA DO FUSTE

Segundo Ahrens e Holbert (1981), uma função de afilamento ou “taper” é uma descrição matemática do perfil longitudinal de um tronco, e seu volume pode ser obtido por meio da integração dessa função, considerando o tronco como sólido de revolução.

Hojer (1903) foi o primeiro a usar funções matemáticas para descrever a forma do fuste e diversos outros vieram em sequência. Sendo assim, muitos modelos de regressão têm sido propostos e utilizados, modelos esses de distintas naturezas matemáticas.

Os modelos de afilamento podem ser classificados de maneira simplificada como modelos de árvore completa ou modelos segmentados (PRODAN et al., 1997), sendo os modelos de árvore completa aqueles que uma única função representa a forma do fuste como um todo, tendo como vantagem a facilidade no ajuste, bem como, geralmente possuem fácil integração para obtenção do volume. Porém é necessário destacar que essas funções possuem forte viés nas estimativas próximo a base e ao ápice do fuste (MAX e BURKHART, 1976; DEMAERSCHALK e KOZAK, 1977; KOZAK, 1988; JIANG et al., 2005; ROJO et al., 2005; BROOKS et al., 2008). Modelos de árvore completa comumente utilizados são os modelos polinomiais, que consistem na comparação entre a variável dependente (d_i/d_{ap}) e os polinômios formados pela variável independente (h_i/ht) (MIGUEL, 2009).

Por outro lado, modelos segmentados são caracterizados pelo uso de submodelos sobrepostos, onde cada um representa uma porção do tronco. A união dos submodelos é feita por meio da imposição de restrições ao modelo geral. Necessário destacar que esses modelos são mais difíceis de serem trabalhados, bem como visam diminuir o viés das estimativas do diâmetro ao longo do fuste através da segmentação em seções do fuste (PRODAN et al., 1997; CAMPOS e LEITE, 2006).

Segundo Machado et al. (2004) os modelos polinomiais são de uso frequente no meio florestal, porém estes muitas vezes não explicam de forma eficiente as deformações existentes na base e no ápice do fuste das árvores. Eisfeld et al. (2004) reafirmam que o uso de modelos polinomiais é vantajoso devido a aliarem eficiência a simplicidade de aplicação, porém salienta também que estes apresentam algumas deficiências na representação da base, podendo comprometer as estimativas.

Uma vez definido o modelo matemático para o afilamento, é possível a determinação do volume entre quaisquer pontos ao longo do fuste, caracterizando uma importante vantagem das funções de afilamento em relação às demais formas de estimar o volume, pois o uso das funções de afilamento possibilita a determinação de diâmetros a qualquer altura, altura comercial relativa a um determinado diâmetro mínimo e a consequente estimativa do volume comercial por sortimento ou total do fuste. Figueiredo (2005) ao comparar equações de volume com as funções de afilamento constatou que a precisão é equivalente, porém, as funções de afilamento são justamente mais interessantes pela estimativa de volume em qualquer porção da árvore, ou seja, para qualquer sortimento de toras.

O modelo de Hradetzky (1976), também conhecido como polinômio de potências fracionárias e inteiras, já foi amplamente aplicado no estudo do afilamento das árvores. É um modelo de grande eficiência e flexibilidade, o qual em análises de comparação de modelos de afilamento já foi considerado por diversos autores como o de melhor desempenho (MACHADO et al., 2004; QUEIROZ et al., 2008; SILVA et al., 2011; KOHLER et al. 2013; TÉO et al., 2013; DAVID et al., 2014).

3.4 INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

O termo Inteligência Artificial (IA), que teve McCulloch e Pitts (1943) como trabalho pioneiro, refere-se à inteligência similar à humana que máquinas artificiais apresentam, tendo, portanto, como objetivo primordial a capacitação do computador a executar funções desempenhadas pelo ser humano usando conhecimento e raciocínio, sendo evidente que a assimilação desse conhecimento é requisito fundamental para a construção desses sistemas inteligentes (MITCHELL, 1997; REZENDE, 2003).

Segundo Russell e Norvig (2010), o trabalho que propiciou uma satisfatória definição operacional da inteligência artificial foi Turing (1950), em “Computação de Máquina e Inteligência”, o qual introduziu o teste de Turing, aprendizado de máquina, algoritmos genéticos e reforço de aprendizagem.

Nesse sentido, o Aprendizado de Máquina (AM) é uma ramificação da IA, em que consiste no processo de indução de uma hipótese ou aproximação de função a partir da experiência passada (FACELI et al., 2011). Conforme Mitchell (1997), AM consiste na capacidade de uma máquina melhorar o desempenho em determinada tarefa por meio de exemplos apresentados, ou seja, a experiência.

AM se baseia no princípio do raciocínio indutivo, ou seja, que informações de um conjunto de exemplos fornecem informação para generalização do todo. Assim, o processo de aprendizado indutivo pode ser dividido em supervisionado, não supervisionado e por reforço (DUDA et al., 2001; REZENDE, 2003; FACELI et al., 2011).

No aprendizado supervisionado, os exemplos fornecidos ao algoritmo contêm a variável resposta desejada, possibilitando assim o ajuste do algoritmo de aprendizado conforme o viés apresentado. Assim, problemas de aprendizado supervisionado podem ser de classificação, no caso de rótulo de dados discreto, ou regressão, em caso de uma variável resposta contínua (REZENDE, 2003; FACELI et al., 2011).

No aprendizado não supervisionado o algoritmo utilizado analisa os exemplos fornecidos quanto a possíveis relações, agrupamentos ou padrões que os dados apresentam. Essa abordagem é altamente utilizada em aplicações de mineração de dados, em que grandes bases de informações são analisadas (WITTEN et al., 2011).

O aprendizado por reforço é caracterizado por um ambiente observável, em que o algoritmo reforça uma ação considerada positiva e pune uma ação considerada negativa (FACELI et al., 2011).

Assim, algoritmos de aprendizado que visam à obtenção de um modelo de regressão ou estimador fazem parte do aprendizado supervisionado, em que, partindo de exemplos da saída desejada é possível a generalização para dados desconhecidos pelo algoritmo treinado.

Tais algoritmos ainda carecem de estudos com aplicações na área florestal, existindo, portanto, uma lacuna no conhecimento do potencial dessas aplicações matemáticas.

3.4.1 *k*-Vizinho mais próximo

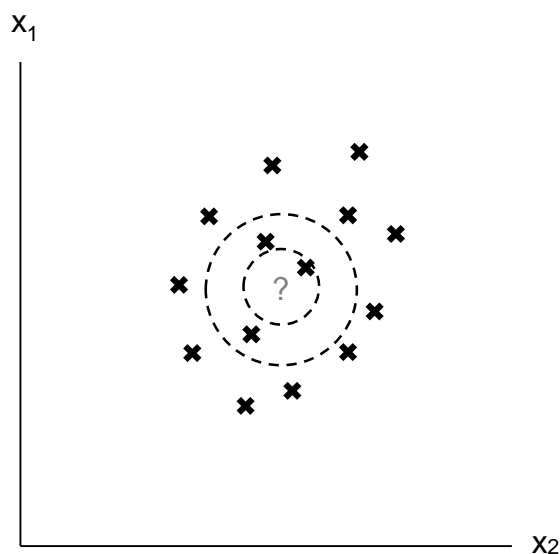
O algoritmo vizinho mais próximo é o mais simples de todos os algoritmos de aprendizado de máquina, baseia-se no pressuposto da estimação de um novo objeto com base nos exemplos do conjunto de treinamento que são próximos a ele com base em uma métrica definida (FACELI et al., 2011). É considerado um algoritmo preguiçoso (*lazy*) no sentido que não há fase de aprendizagem com a construção de modelo e posterior generalização para todos os dados, mas sim a memorização dos dados de treinamento (FACELI et al., 2011; WITTEN et al., 2011).

Inicialmente aplicado por Fix e Hodges (1951)¹, posteriormente teorizado por Cover e Hart (1967), o algoritmo NN (*Nearest neighbor*) procura classificar ou estimar determinado atributo com base em seu vizinho mais próximo. O algoritmo *k*-NN é uma variação simples do algoritmo original, em que considera a distância dos *k* vizinhos mais próximos (DUDA et al., 2001).

A FIGURA 1 apresenta um exemplo ilustrativo, considerando a utilização de um vizinho próximo ou 3 vizinhos próximos, tendo como base duas variáveis explicativas. Assim é possível visualizar a influência do número de vizinhos a serem considerados, pois essa reflete diretamente na estimativa da variável de interesse.

¹ Publicado oficialmente em: FIX, E.; HODGES, J. L. Discriminatory analysis – Nonparametric discrimination: Consistency properties. **Internat. Statist. Rev.** 57 238–247, 1989.

FIGURA 1 – IMPACTO DO VALOR DE K NO ALGORITMO K-NN



FONTE: Adaptado de DUDA et al. (2001).

Quanto a aplicações para fins florestais, *k*-NN tem sido amplamente utilizado em sensoriamento remoto para generalização de atributos da floresta com base em dados de inventário florestal (MCROBERTS et al., 2002; MCROBERTS et al., 2007; KOUKAL et al., 2007; KAJISA et al., 2008; CHIRICI et al., 2008; BAFFETTA et al., 2009; FUCHS et al., 2009; MCINERNEY e NIEUWENHUIS, 2009; NOTHDURFT et al., 2009; BREIDENBACH et al., 2010; MCROBERTS 2012; OHMANN et al., 2012; WILSON et al., 2012; OHMANN et al., 2014; MANSUY et al., 2014; BLANCHETTE et al., 2015; MCROBERTS et al., 2015).

A aplicação direta na regressão para fins de mensuração e inventário florestal ainda é muito restrita, com trabalhos relativamente antigos, que por outro lado não popularizaram a técnica. Maltamo e Kangas (1998) aplicaram para modelagem da distribuição da área basal a fim de comparar a exatidão das estimativas obtidas pelo *k*-NN com estimativas da estrutura e volume do povoamento, bem como, com previsões de crescimento. Concluíram por fim que os métodos baseados na regressão *k*-NN provaram dar uma descrição mais precisa do povoamento do que os métodos paramétricos.

Sironen et al. (2003) compararam distintos métodos não paramétricos, tais como: vizinho mais próximo, vizinho mais similar e modelos aditivos generalizados, a fim de obter modelos de crescimento individual e, por fim, concluíram que dentre os três métodos, *k*-NN se mostrou mais robusto.

Fehrmann et al. (2008) compararam *k*-NN com modelagem mista para estimativa de biomassa individual, e relatam que o método *k*-NN obteve resultados similares, porém mais acurados que a modelagem convencional. Eskelson et al. (2009) reportaram *k*-NN como uma importante ferramenta a fim de estimativas de valores ausentes em bancos de dados de inventário e monitoramento da floresta.

Haara e Kangas (2012) fazem uma interessante abordagem quanto a comparação entre *k*-NN e a regressão comumente aplicada a problemas florestais. Os autores destacam que a análise de regressão paramétrica tem a vantagem da teoria estatística bem conhecida, ao passo que as propriedades estatísticas de *k*-NN são menos estudadas. Com base em simulações simples, destacam ainda que as estimativas são muito similares quanto ao erro nas estimativas em uma base de dados equilibrada e homogênea, porém à medida que o conjunto de dados era desequilibrado e a não linearidade acentuada, as diferenças entre as estimativas aumentaram, com o *k*-NN apresentando resultados menos tendenciosos que a regressão convencional.

Sanquetta et al. (2013) analisaram o desempenho de distintos métodos para a estimativa do carbono fixado em indivíduos de *Araucaria angustifolia* Bert. O. Ktze., comparando *k*-NN e modelos alométricos² de regressão linear, método comumente aplicado para esse fim. Os autores concluem que os resultados obtidos não diferem estatisticamente, porém destacam que o método *k*-NN não pressupõe qualquer forma específica da função que relaciona o carbono e os parâmetros biométricos da árvore, visto que a escolha de tal função é difícil de encontrar. O vizinho mais próximo pode ser similar aos métodos convencionais quanto à precisão, mas também pode simplesmente livrar das perguntas e testes para a escolha das equações alométricas dentre diversas relações possíveis.

Wojciechowski (2015) apresentou uma ferramenta de *software on line* para auxiliar nas estimativas de volume, biomassa e principalmente Carbono, chamada de JCarbon. A ferramenta utiliza além da técnica usual de regressão, a possibilidade de obter estimativas via mineração de dados, no caso especificamente com o algoritmo *k*-NN.

² Alometria: alteração na parte modificando o todo. (GAYON, J. History of the Concept of Allometry. American Zoologist. v. 40. p.748-758, 2000)

Sanquetta et al. (2015a) ao compararem k -NN e modelos alométricos para estimativas de biomassa, consideraram os modelos k -NN são comparáveis e em alguns casos mais precisos que modelos alométricos tradicionais, a exemplo do modelo de Schumacher-Hall. Destacam ainda que as principais virtudes da técnica k -NN estão na versatilidade, flexibilidade e simplicidade, além da não exigência de satisfazer os pressupostos de regressão.

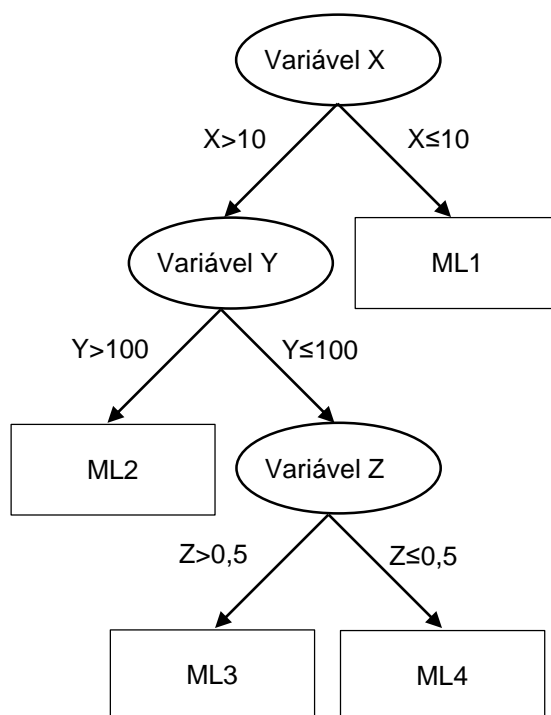
Por outro lado, Mognon et al. (2014) ao compararem a técnica de vizinhos mais próximos e regressão linear para estimativa da biomassa de colmos de bambu, relataram a maior precisão dos modelos de regressão. Os autores atribuem o pior desempenho da técnica de aprendizado de máquina à limitação da base de dados, composta apenas por 25 colmos para ajuste e 13 para validação. Sendo assim, afirmam que o efeito do tamanho da amostra é determinante para a obtenção de estimativas, porém ainda mais para o algoritmo k -NN. Portanto, para os autores, quando não existe disponível uma grande base de dados, a regressão linear deve ser preferida.

3.4.2 *Random Forest*

Quinlan (1992) desenvolveu o algoritmo de árvores de decisão/regressão, o qual se baseia na ideia simples da construção de uma árvore de regressão, dividindo as variáveis explicativas conforme suas semelhanças e dissimilaridades (WANG e WITTEN, 1997).

Desenvolvido por Breiman (2001), o algoritmo *Random Forest* (RF) é um desenvolvimento direto do algoritmo árvore de decisão, tornando assim possível reduzir a variabilidade em cada nó e obter estimativas via um modelo linear local. A FIGURA 2 apresenta um exemplo ilustrativo do funcionamento de uma árvore de regressão, em que determinado atributo é explicado por três variáveis (X, Y e Z) e sua estimativa é obtida por meio de quatro modelos lineares (ML1, 2, 3 e 4). Destaque a divisão binária das características dos atributos em cada nó.

FIGURA 2 – REPRESENTAÇÃO DE UMA ÁRVORE DE REGRESSÃO



FONTE: Adaptado de HALL et al. (2009).

O algoritmo RF consiste em um conjunto de árvores de regressão treinadas aleatoriamente, ou seja, é um conjunto de árvores treinadas em que cada árvore dá um voto que indica sua decisão sobre o valor do objeto.

Segundo Seyedhosseini e Tasdizen (2015), o algoritmo RF é um classificador/regressor cuja ideia principal da “Floresta aleatória” é fazer crescer um conjunto de árvores de decisão tal que a correlação entre essas árvores permaneça tão baixa quanto possível. Tal condição é alcançada através do método *bagging*, em que a aleatoriedade é injetada na floresta usando um subconjunto diferente de amostras de treinamento para cada árvore, assim, randomicamente são extraídas m instâncias do conjunto de treinamento, visando treinar determinado número de árvores da floresta (valor esse definido pelo usuário). Além disso, em cada nó da árvore um subconjunto aleatório das variáveis de entrada é usado para aprender a função de partição. A ideia principal é fazer com que as árvores de decisão sejam as mais independentes quanto possível, melhorando a robustez e a generalização do conjunto (BREIMAN, 2001).

Árvores de decisão podem superar as abordagens de regressão clássicas, especialmente quando ocorrem relações complexas e não lineares entre as variáveis, bem como os resultados das árvores de decisão são mais fáceis de interpretar do que outros modelos de regressão (STROBL et al., 2009). Porém Strobl et al. (2008) destacam que a principal fraqueza de modelos de árvore simples é a instabilidade frente a pequenas alterações nos dados de aprendizagem. Nesse sentido o algoritmo *Random Forest* resolve o problema dessa instabilidade através da média de um conjunto de árvores em um modelo complexo mais robusto (BREIMAN, 2001; CUTLER et al., 2007; STROBL et al., 2009).

Moreno-Fernández et al. (2015) ao analisarem procedimentos paramétricos e não paramétricos para modelar a regeneração natural, incluindo o algoritmo RF, destacam que para esse tipo de análise, visto que dados de regeneração raramente satisfazem as premissas básicas para regressão, tais como normalidade dos resíduos, linearidade dos parâmetros, homogeneidade da variância e independência entre preditores, métodos não paramétricos apresentam grande potencial para utilização, pois assim é possível modelar essa estrutura complexa. Por outro lado, os autores destacam como limitação não ser possível a interpretação de parâmetros como em um modelo tradicional.

Cutler et al. (2007) destacam procedimentos altamente computacionais tais como o RF não apresentam relações simples entre variáveis preditoras e variáveis resposta, tais como uma equação matemática ou representação gráfica como árvores de regressão, sendo que a falta dessa representação simples pode tornar sua interpretação difícil. Os autores destacam também que o RF não é uma ferramenta para inferência estatística tradicional, não sendo, portanto, adequado para a ANOVA e testes de hipóteses. Ele não computa p valores, coeficientes de regressão, nem intervalos de confiança, mas destacam como vantagem o procedimento não necessitar de suposições acerca da distribuição dos dados, assim como alta capacidade de modelar interações complexas entre um grande número de variáveis preditivas.

De maneira geral, o algoritmo RF ainda foi pouco aplicado, mas é de interesse crescente dado o bom desempenho prático. Por outro lado, no aspecto teórico o seu poder preditivo continua em grande parte inexplicável (SCORNET, 2015).

Johansson et al. (2014) ao compararem o desempenho do RF em relação ao *k*-vizinho mais próximo e redes neurais artificiais, utilizando para tal 33 bases de dados distintas, concluíram que as previsões obtidas pelo RF são significativamente mais eficientes que as supracitadas para a maior parte dos estudos de caso.

A utilização do algoritmo *Random Forest* para fins florestais é praticamente nula. Assim como o algoritmo *k*-vizinho mais próximo, tem sido mais utilizado no sensoriamento remoto (PIERCE et al., 2012; GRINAND et al., 2013; BEIJMA et al., 2014; CHEN et al., 2014; GORGENS et al., 2015; TATSUMI et al., 2015; TRAMONTANA et al., 2015; WERE et al., 2015). Prasad et al. (2006) ao analisarem RF para mapeamento da ocorrência de 4 espécies florestais destacam que esse tipo de técnica pode ser utilizado para extrapolação de qualquer variável amostrada para o todo, obtendo um maior nível de confiança do que com outros métodos convencionais na predição para mapeamento.

Lawler et al. (2006) fizeram uma interessante comparação no desempenho de seis métodos estimativos de ocorrência de espécies frente a mudanças climáticas, em que dentre esses métodos testaram RF e RNAs, concluíram pôr fim a superioridade do RF em relação aos demais métodos, destacando a taxa de acerto do RF maior que 99% e dos demais métodos cerca de 86%.

3.4.3 Redes Neurais Artificiais

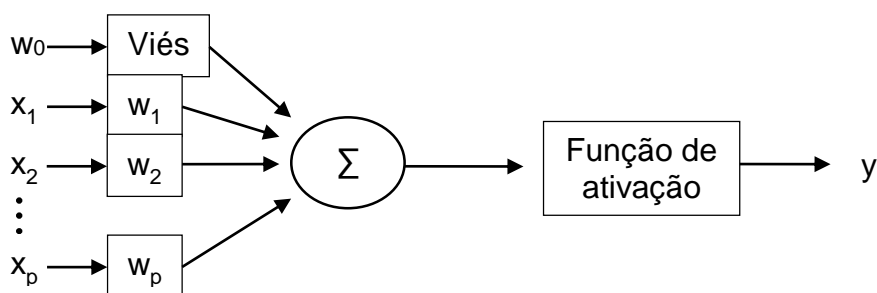
As redes neurais artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso humano, os quais adquirem conhecimento através de um processo de aprendizagem. Conforme Haykin (2001), uma RNA é um processador composto por diversas unidades de processamento simples, paralelamente distribuídos e densamente interligados, os neurônios, que têm como finalidade o armazenamento de conhecimento e informações a respeito de determinada tarefa à qual a rede se propõe. Deste modo, a semelhança quanto ao cérebro humano se dá quanto ao conhecimento adquirido pela rede a partir de um processo de aprendizagem, bem como quanto as forças de conexão entre os neurônios, conhecidas como pesos sinápticos, que são utilizadas para armazenar o conhecimento.

Assim, os neurônios são dispostos em uma ou mais camadas interligadas por um grande número de conexões, em um arranjo semelhante aos das redes

neurais biológicas. Dado um conjunto de dados, as RNA têm a capacidade de aprendizado de informações obtidas pelo conjunto, visando posteriormente à generalização do conhecimento assimilado para um conjunto de dados desconhecidos. O procedimento utilizado nesse processo de aprendizado é chamado de algoritmo de aprendizado (HAYKIN, 2001), e tem como função modificar os pesos sinápticos a fim de alcançar o objetivo ao qual a rede foi planejada. As RNAs possuem ainda a capacidade de extrair características não explícitas de um conjunto de informações que lhes são fornecidas (DIAMANTOPOULOU, 2005; BRAGA et al., 2007; GORGENS et al., 2009; BINOTI et al., 2014b). Assim sendo, uma rede neural é caracterizada pelo seu padrão de conexões entre os neurônios e número desses, ou seja, sua arquitetura; seu método de determinação dos pesos das conexões, ou seja, seu algoritmo de aprendizagem (FACELI et al., 2011; ÖZÇELIK et al., 2013).

Conforme anteriormente mencionado, o neurônio é a unidade de processamento fundamental de uma RNA (HAYKIN, 2001), assim sendo, a FIGURA 3 ilustra a estrutura de um neurônio simples em que é possível identificar o conjunto de pesos sinápticos (w_1, \dots, w_p) associados as variáveis (x_1, \dots, x_p), em seguida o somador dos sinais de entrada ponderados pelos respectivos pesos sinápticos e uma função de ativação, para restringir a amplitude de saída do neurônio.

FIGURA 3 – MODELO SIMPLIFICADO DE UM NEURÔNIO E SUA ESTRUTURA

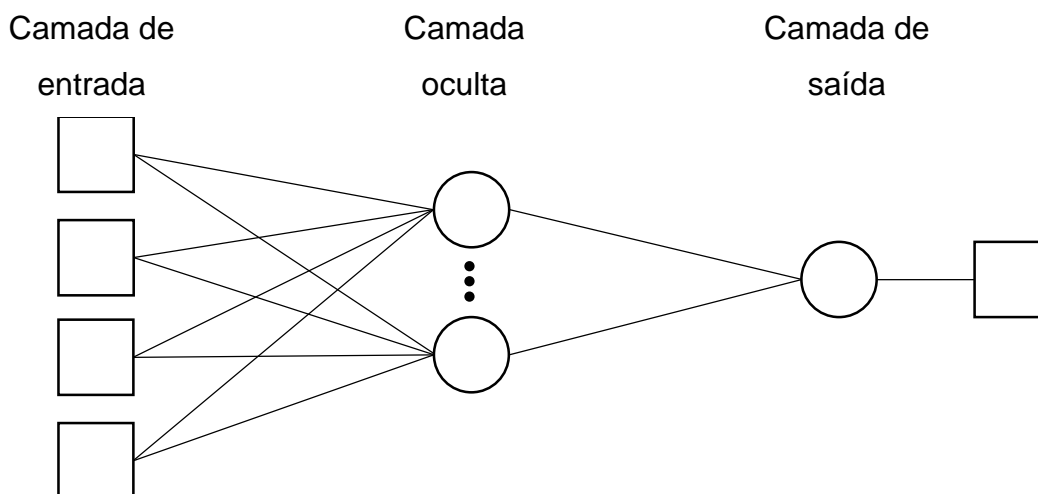


FONTE: Adaptado de HAYKIN (2001)

Foram utilizadas RNAs do tipo perceptrons de múltiplas camadas (MLP – *multilayer perceptrons*), visto que possuem a capacidade universal de aproximação de qualquer função contínua (CYBENKO, 1989; HORNIK et al., 1989). Segundo Haykin (2001), as redes MLP são tipicamente definidas pela camada de entrada, uma ou mais camadas ocultas constituídas por neurônios e uma camada de saída

(FIGURA 4). Redes do tipo MLP pertencem a classe das redes do tipo *feedforward*, visto que essas redes não contêm épocas e as saídas dependem unicamente das instâncias de entrada (CYBENKO, 1989; WITTEN et al., 2011).

FIGURA 4 – REPRESENTAÇÃO GRÁFICA DA ARQUITETURA DE UMA RNA



FONTE: Adaptado de HALL et al. (2009)

O algoritmo *backpropagation*, ou retropropagação de erro, desenvolvido por Rumelhart et al. (1986), tem como característica a otimização baseada no gradiente descendente da função de erro, em que baseado em regras de aprendizagem é possível a obtenção do melhor ajuste, avaliando o erro e tentando reajustá-lo.

Conforme Haykin (2001), o processamento do algoritmo *backpropagation* distingue-se em dois passos. O primeiro conhecido como passo para frente, ou propagação, em que os pesos sinápticos iniciais são aleatoriamente definidos e se mantêm inalterados na época. O segundo passo, conhecido como retropropagação, é caracterizado como o movimento iniciado na camada de saída retornando às camadas ocultas, em que é calculado o gradiente local de cada neurônio em função do erro, reajustando, portanto, o peso sináptico de cada neurônio, a cada época. Esse processo é repetido até que o critério de convergência escolhido seja alcançado, sendo geralmente esse critério a minimização do erro quadrático ou número máximo de épocas³.

³ Também conhecido como ciclo.

Segundo Haykin (2001) a taxa de aprendizagem corresponde a uma constante para a trajetória dos cálculos dos pesos sinápticos, ou seja, determinam o tamanho do passo que a rede realizará a cada época no ajuste dos pesos. Portanto, quanto menor a taxa de aprendizagem, menor serão as oscilações dos pesos sinápticos da rede de uma interação para outra, tornando a convergência muito lenta. Em contrapartida ao utilizar uma taxa de aprendizagem muito alta, as grandes oscilações nos pesos podem tornar a rede instável, bem como podem não alcançar a minimização do erro conforme propõe o algoritmo.

Visando aumentar a taxa de aprendizagem, e consequentemente acelerar a convergência da rede, Rumelhart et al. (1986) propuseram a inclusão da constante *momentum*, que tem por finalidade o controle da realimentação dos pesos sinápticos a cada época. Conforme Haykin (2001), a incorporação da constante *momentum* no algoritmo representa pequena variação no ajuste dos pesos, porém pode apresentar efeitos benéficos na aprendizagem da rede, bem como pode evitar que o processo de aprendizado termine em um mínimo local⁴.

Sendo assim, para o algoritmo *backpropagation*, além das características já referidas das RNAs, como arquitetura, função de ativação e seus pesos interligados, os termos *momentum* e taxa de aprendizado compõem parte importante na formação e ajuste da rede, bem como o número de épocas de treinamento (ATTOH-OKINE, 1999).

No tocante à aplicação em regressão para fins florestais, as RNAs podem ser consideradas mais difundidas que os algoritmos anteriores, mas ainda apresentam utilização muito recente. Dentre sua utilização como aproximação de função, destacam-se estudos para estimativas do volume individual (SILVA et al., 2009; GORGENS et al., 2009; BLANCO et al., 2010; DIAMANTOPOULOU e MILIOS, 2010; ÖZÇELIK et al., 2010; BINOTI et al., 2014a, b; PELLI, 2013; CORDEIRO et al., 2015), altura (BRANDÃO, 2007; DIAMANTOPOULOU e ÖZÇELIK, 2012; BINOTI et al., 2013; CASTAÑO-SANTAMARÍA et al., 2013; ÖZÇELIK et al., 2013; PELLI, 2013), diâmetro relativo e estudo da forma (PEREIRA, 2008; LEITE et al., 2011; SOARES et al., 2011; SOARES et al., 2013; SOUZA, 2013;

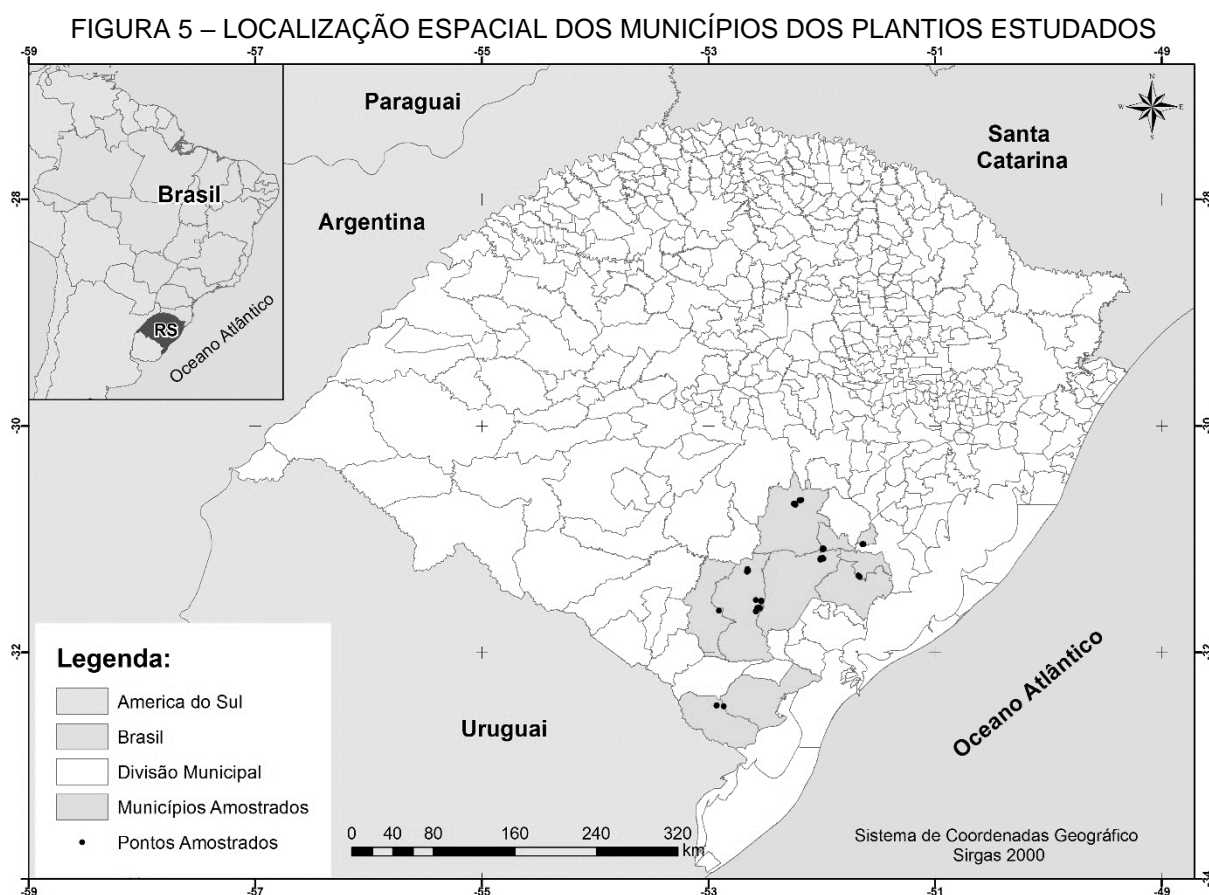
⁴ O ajuste do algoritmo *backpropagation* é baseado no gradiente descendente. Assim há risco de o algoritmo ficar preso em mínimos locais onde qualquer pequena mudança nos pesos sinápticos acarreta um aumento do erro, portanto o algoritmo entende que esse é o momento de parada do treinamento. Entretanto, em algum lugar do espaço de busca do algoritmo há um conjunto de pesos sinápticos para o qual o erro é menor do que o mínimo local, ou seja, o mínimo global (HAYKIN, 2001).

ÖZÇELİK et al., 2014; SANTOS, 2014; ARAUJO, 2015; SCHIKOWSKI et al., 2015), distribuição de diâmetros (LEDUC et al., 2001; CAI et al., 2010; DIAMANTOPOULOU et al., 2015), volume da casca (DIAMANTOPOULOU, 2005).

4 MATERIAL E MÉTODOS

4.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO

A presente pesquisa foi realizada tendo como base dados de plantios da espécie *Acacia mearnsii* De Wild., popularmente conhecida como acácia-negra. Os plantios estão situados no estado do Rio Grande do Sul, divididos em três regiões: Cristal, Encruzilhada do Sul e Piratini. A região de Cristal é composta pelos municípios de Canguçu, Chuvisca, Cristal, São Lourenço do Sul; região de Piratini é composta pelos municípios de Arroio Grande, Jaguarão, Pinheiro Machado, Piratini; região de Encruzilhada do Sul é composta pelos municípios de Amaral Ferrador e Encruzilhada do Sul (FIGURA 5).



FONTE: A autora (2016).

O clima na região é subtropical úmido, do tipo Cfa de Köppen (MORENO, 1961). Tem como característica chuvas distribuídas ao longo do ano, porém com

maior concentração no verão. Temperatura média no mês mais frio inferior a 18°C e temperatura média superior a 22°C no mês mais quente, tendo portando verões quentes e geadas pouco frequentes (BRASIL, 1992; CARVALHO, 2003).

Conforme Mochiutti (2007), em estudo para os mesmos locais dos plantios, os solos de Cristal e Piratini foram classificados como NEOSSOLO REGOLÍTICO distrófico e NEOSSOLO LITÓLICO distrófico, respectivamente. Já a região de Encruzilhada do Sul como um todo tem a predominância de ARGISSOLOS VERMELHO-AMARELOS nas áreas de planalto, bem como ARGISSOLOS VERMELHOS Distróficos nas áreas onduladas (CUNHA et al., 2005). A nomenclatura dos autores foi atualizada conforme EMBRAPA (2006).

4.2 CUBAGEM

Em cada região foram amostrados quatro povoamentos de forma a abranger toda a rotação da cultura (10 anos). Em Cristal foram amostrados povoamentos com 1, 2, 5 e 10 anos, em Encruzilhada do Sul com 1, 3, 5 e 10 anos e em Piratini com 1, 2, 5 e 9 anos.

Em cada povoamento foram instaladas de forma aleatória quatro parcelas circulares com 10 m de diâmetro, em que as árvores tiveram seu diâmetro a 1,30 m (dap) medido e posteriormente foram derrubadas para a realização da cubagem. Os indivíduos mortos foram desconsiderados.

A cubagem rigorosa foi realizada pelo método de Huber, com a mensuração dos diâmetros relativos aos 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85% e 95% da altura total do fuste, totalizando 10 seções. A base de dados foi dividida em 60% para treinamento dos algoritmos e os 40% restantes para a validação dos modelos. A TABELA 1 apresenta a distribuição dos indivíduos amostrados em relação a idade e classe de dap, totalizando 683 indivíduos cubados. As medições foram realizadas com a utilização de fita métrica.

TABELA 1 – DISTRIBUIÇÃO DA FREQUÊNCIA DOS INDIVÍDUOS AMOSTRADOS POR CLASSES IDADE E DE DIÂMETRO EM FUNÇÃO DOS GRUPOS DE AJUSTE E VALIDAÇÃO

Conj.	Idade (anos)	Classe de dap (cm)			Total
		> 5	5 - 9,9	≥ 10	
Treino	1	54	58	2	114
	2	1	31	26	58

	3	3	35	6	44
	5	3	21	62	86
	9	-	9	27	36
	10	-	18	55	73
Validação	1	30	42	1	73
	2	1	21	16	38
	3	2	16	6	24
	5	5	20	43	68
	9	-	2	15	17
	10	1	14	37	52
Total Geral		100	287	296	683

FONTE: A autora (2016).

4.3 EQUAÇÃO DE VOLUME

O modelo de Schumacher e Hall (1933), em sua forma linearizada, é apresentado na Equação 1.

$$\ln(v) = \beta_0 + \beta_1 \ln(dap) + \beta_2 \ln(ht) + \varepsilon_i \quad (1)$$

Em que:

v = volume do fuste (m^3);

dap = o diâmetro a 1,30 m do solo (cm);

ht = altura total (m);

\ln = logaritmo neperiano;

β_n = coeficientes do modelo;

ε_i = erro aleatório inerente à regressão.

O ajuste foi realizado com o procedimento *model* e *reg* (SAS, 2002), a fim da confirmação do atendimento das condicionantes da regressão: normalidade (teste de Shapiro-Wilk), homogeneidade (teste de White) e independência dos resíduos (Durbin-Watson). Optou-se pela estratificação dos dados com base na classe de diâmetro, dividindo em 3 classes: abaixo de 5 cm de dap ; de 5 a 10 cm e acima de 10 cm.

4.4 FUNÇÃO DE AFILAMENTO

O modelo conforme descrito por Hradetzky (1976) é apresentado na Equação 2.

$$\frac{d_i}{dap} = \beta_0 + \beta_1 \left(\frac{h_i}{ht}\right)^{p1} + \beta_2 \left(\frac{h_i}{ht}\right)^{p2} + \dots + \beta_n \left(\frac{h_i}{ht}\right)^{pn} + \varepsilon_i \quad (2)$$

Em que:

d_i = diâmetro i ao longo do fuste (cm);

dap = o diâmetro a 1,30 m do solo (cm);

h_i = altura i ao longo do fuste (m);

ht = altura total (m);

β_n = coeficientes do modelo;

p_n = potências selecionadas para compor o modelo;

ε_i = erro aleatório inerente à regressão.

As potências foram selecionadas pelo procedimento *stepwise*, considerando um nível de significância de 15% ($\alpha = 0,15$) no teste F de cada coeficiente parcial do modelo, utilizando o *Proc reg* do *software* SAS (2002). Os expoentes utilizados foram 0,005; 0,01; 0,02; 0,03; 0,04; 0,05; 0,06; 0,07; 0,08; 0,09; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1; 2; 3; 4; 5; 10; 15; 20 e 25, conforme também aplicado por Téó et al. (2013). Adicionalmente foi aplicado o procedimento *model* (SAS, 2002), com base no modelo com as potências selecionadas, a fim da confirmação do atendimento das condicionantes da regressão: normalidade (teste de Shapiro-Wilk), homogeneidade (teste de White) e independência dos resíduos (Durbin-Watson). A estimativa do volume foi obtida pela fórmula $PoliHradetzkyVi$, integrada ao Florexel (ARCE et al., 2000).

Diversos autores afirmam que a estratificação da base de dados por classe de diâmetro ou idade converge em melhores estimativas para o afilamento (MACHADO et al., 2004; MULLER, 2004; QUEIROZ et al., 2008; SOUZA, 2009; KOHLER et al. 2013), sendo assim, optou-se pela estratificação com base na classe de diâmetro, dividindo a base de dados em 3 classes: abaixo de 5 cm de dap ; de 5 a 10 cm e acima de 10 cm.

4.5 ALGORITMOS DE APRENDIZADO DE MÁQUINA

Foram avaliados três algoritmos de aprendizado de máquina: *k*-vizinho mais próximo (*k*-NN); algoritmo *Random Forest* (RF); e Redes neurais artificiais (RNA).

A base de dados foi dividida em 60% para treinamento dos algoritmos e os 40% restantes para a validação dos modelos, podendo assim avaliar a capacidade de generalização dos mesmos. Essa divisão foi realizada respeitando a proporção de indivíduos por classe de idade, bem como, também foi adotada nos ajustes dos modelos de regressão.

Todos os algoritmos supracitados foram treinados no *software* Weka 3.7.12 (HALL et al., 2009), utilizando a opção de teste de validação cruzada em 10 grupos, o qual consiste na divisão do conjunto de teste em 10 grupos de tamanho aproximadamente igual. Conforme Faceli et al. (2011) consiste na utilização de, no caso do presente trabalho, 9 grupos utilizados no treinamento do preditor para posterior teste com o grupo restante. Esse procedimento foi repetido 10 vezes, utilizando uma partição diferente em cada época e com ajuste do modelo ao decorrer desses testes.

Foram adotadas como variáveis explicativas para a estimativa de diâmetro ao longo do fuste a idade, o dap, a altura total e a altura relativa.

Os modelos de AM foram treinados para obtenção do volume também com as variáveis idade, dap, altura total e altura relativa, porém foi utilizada uma lógica do volume acumulado a dada altura relativa, conforme a Equação 3.

$$v_i = \sum_{i=0}^{Nt} g_i L_i \quad (3)$$

Em que:

v_i = volume do fuste na altura i (m^3);

g_i = área transversal dos toretes medidos (m^2);

L_i = comprimento dos toretes medido na cubagem (m);

Nt = número de toretes.

4.5.1 *k*-Vizinho mais próximo

Os *k*-vizinhos mais próximos de cada instância são definidos conforme a métrica selecionada, como observado na Equação 4 foi utilizada a distância Euclidiana, a qual é comumente utilizada para aplicação do algoritmo, ainda que haja outras opções no Weka, tais como: distância de Chebyshev, Manhattan e Minkowski (WITTEN et al., 2011). Visando eliminar a influência da escala das variáveis, estas são normalizadas ou padronizadas⁵ (MITCHELL, 1997; FACELI et al., 2011), sendo nesse caso normalizadas conforme a Equação 5.

$$d(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

Em que:

d =distância Euclidiana entre dois pontos $P(p_1, p_2, \dots, p_n)$ e $Q(q_1, q_2, \dots, q_n)$;

P e Q = diferentes indivíduos com n atributos ou variáveis explicativas.

$$a_i = \frac{v_i - \min v_i}{(\max v_i - \min v_i)} \quad (5)$$

Em que:

a_i = instância normalizada;

v_i = valor atual da instância i a ser normalizada;

$\max v_i$ e $\min v_i$ = valores máximo e mínimos que o atributo i apresenta.

Aha (1992) introduziu o conceito da ponderação dos atributos, evitando assim o viés na estimativa que dados com ruído podem induzir, tornando assim a aprendizagem por *k*-vizinhos mais próximos mais robusta. Assim sendo, a instância a ser estimada é obtida de acordo com a sua distância em relação aos exemplos de teste (EQUAÇÃO 6).

⁵ Padronizar – subtrair a média e dividir pelo desvio. Resulta em distribuição com 0 de média e 1 de desvio. Normalizar – dividir pelo valor máximo ou subtrair o valor mínimo e depois divide pelo intervalo ($\max - \min$) WITTEN et al. (2011)

$$\hat{f}(x_q) = \frac{\sum_{i=1}^k w_i * f(x_i)}{\sum_{i=1}^k w_i} \quad (6)$$

Em que:

$\hat{f}(x_q)$ = instância de valor desconhecido a ser estimado;

$f(x_i)$ = instâncias observadas utilizadas como base para a estimativa;

w_k = fator de ponderação ou peso relativo ao vizinho k ;

k = número de vizinhos utilizados na predição.

Na presente pesquisa padronizou-se a ponderação dos k -vizinhos mais próximos como o inverso de suas respectivas distâncias ($1/d$). Método esse recomendado por Bradzil et al. (2003) e Sanquetta et al. (2013).

Sendo assim, as configurações de entrada para esse algoritmo no *software* Weka foram: 20 vizinhos próximos com validação cruzada ativada, essa opção permite localizar k ótimo de vizinhos próximos tendo como base o erro médio absoluto (WITTEN et al., 2011), ponderação das distâncias por $1/d$ e procura linear pelo vizinho próximo utilizando a distância Euclidiana.

4.5.2 *Random Forest*

O algoritmo RF permite limitada customização quanto aos seus parâmetros de entrada. Foram mantidas as configurações *default* do programa Weka, com exceção do número de árvores a serem construídas, pois o programa sugere a construção de 100 árvores de regressão, porém Grimm et al. (2008) e Were et al. (2015) sugerem que números mais elevados fornecem resultados mais estáveis. Sendo assim foram construídas 1.000 árvores de regressão a fim de melhorar a robustez do algoritmo.

4.5.3 Redes neurais artificiais

O *software* Weka utiliza como função de ativação a função logística (Equação 7), do tipo sigmoidal, variando de 0 a 1, sendo, portanto, requerido que os dados sejam normalizados a essa escala.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

Para uma aproximação de função simulada, Refenes et al. (1994) relataram que para uma RNA de uma e duas camadas, uma taxa de aprendizagem de 0,2 e um termo *momentum* entre 0,3 e 0,5 produzem a melhor combinação de convergência, sendo utilizado na presente pesquisa o valor de 0,4 para o termo *momentum*.

Sendo assim, para o treinamento da RNA foram fornecidos os seguintes parâmetros de entrada no Weka: *momentum* 0,4 e número de épocas 1000, todos como termos fixos.

Quanto a taxa de aprendizagem e o número de neurônios na camada oculta, estes foram obtidos utilizando a ferramenta *CVParameterSelection* disponível no Weka, em que é possível fazer a otimização de parâmetros dos algoritmos. É necessário destacar que é um procedimento computacionalmente pesado cujo processamento pode levar dias, por esse motivo adotou-se valores de literatura para alguns desses parâmetros.

Foi utilizada apenas uma camada oculta, pois conforme o “Teorema da aproximação universal”, apenas uma camada oculta já é suficiente para uma rede MLP realizar a aproximação de qualquer função contínua (CYBENKO, 1989; HAYKIN, 2001).

4.6 AVALIAÇÃO DOS MODELOS

Os modelos foram ajustados e avaliados quanto à estimativa do diâmetro ao longo do fuste e volume total.

Os resultados obtidos pelos modelos foram avaliados conforme a correlação de Pearson (r) entre os valores observados e preditos (Equação 8), raiz quadrada do erro médio (RSME%) em porcentagem (Equação 9), bem como, através da análise gráfica dos resíduos absolutos e histograma de frequência dos erros percentuais.

$$r = \frac{\text{Cov}(y_i, \hat{y}_i)}{\sqrt{(\text{Var}(y_i) * \text{Var}(\hat{y}_i))}} \quad (8)$$

Em que:

r = coeficiente de correlação de Pearson;

$\text{Cov}(y_i, \hat{y}_i)$ = covariância entre os valores observados e preditos;

$\text{Var}(y_i)$ e $\text{Var}(\hat{y}_i)$ = variância dos valores observados e preditos, respectivamente.

$$\text{RSME}_{\%} = \frac{100}{\bar{y}_{\text{real}}} \sqrt{\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \right)} \quad (9)$$

Em que:

$\text{RSME}_{\%}$ = raiz quadrada do erro médio;

\bar{y}_{real} = média dos valores observados;

y_i = valor observado na observação i (d_i ou v_i);

\hat{y}_i = valor estimado para a observação i ;

n = número de observações.

Complementando o teste de acurácia dos modelos de afilamento, foram utilizadas estatísticas adicionais que se baseiam em testes nos resíduos (Equações 10, 11, 12, 13 e 14), conforme metodologia indicada por Parresol et al. (1987) e Figueiredo Filho et al. (1996).

$$\text{Desvio Médio (D)} = \left(\sum \text{Diff}_i \right) / n \quad (10)$$

$$\text{Desvio Médio Absoluto (|D|)} = \left(\sum |\text{Diff}_i| \right) / n \quad (11)$$

$$\text{Desvio Padrão das Diferenças (SD)} = \frac{\sqrt{\sum (\text{Diff}_i^2) - \frac{(\sum \text{Diff}_i)^2}{n}}}{(n - 1)} \quad (12)$$

$$\text{Soma do Quadrado do Resíduo Relativo (SQRR)} = \sum (\text{Diff}_i / y_i)^2 \quad (13)$$

$$\text{Porcentagem dos Resíduos (RP)} = \sum (\text{Diff}_i / y_i) * 100 / n \quad (14)$$

Em que:

Diff_i = desvio das variáveis observadas e estimadas ($y_i - \hat{y}_i$);

y_i = valor observado na observação i (d_i ou v_i);

\hat{y}_i = valor estimado para a observação i ;

n = número de observações.

O desvio (D) indica a existência ou não de tendências entre os resíduos, porém não mede o quanto os valores residuais estão afastados do eixo zero. Já a porcentagem dos resíduos (RP) mostra a amplitude dos erros, o SSRR relaciona o tamanho de cada resíduo em relação ao seu valor real, e o desvio padrão das diferenças (SD) mostra a homogeneidade entre os resíduos (PARRESOL et al., 1987; SOUZA, 2009).

Com base nas sete estatísticas supracitadas, os modelos foram ordenados a fim de definir qual obteve o melhor desempenho em média. Esta análise foi feita por atribuição de notas, sendo a menor nota dada ao melhor modelo da referida estatística. Portanto, o que apresentar o menor somatório será considerado o de melhor desempenho.

Foi realizada a análise gráfica dos resíduos quanto a distribuição dos erros absolutos e histograma dos erros percentuais distribuídos em classes, similar a Binoti et al. (2014a, b). Essa análise tem por objetivo ratificar a decisão do ranqueamento, bem como, identificar possíveis tendências ao longo da linha estimativa.

5 RESULTADOS E DISCUSSÃO

5.1 IA x FUNÇÃO DE VOLUME

5.1.1 Ajuste e teste das condicionantes de regressão do modelo de Schumacher e Hall

A TABELA 2 apresenta os coeficientes ajustados para o modelo de Schumacher e Hall, sendo todos os coeficientes significativos. Os indicadores de ajuste como a correlação entre valores observados e preditos e RSME% demonstram que os modelos são adequados para o uso na estimativa do volume. Indivíduos abaixo de 5 cm de dap não foram considerados na análise em virtude da limitação do programa Weka no ajuste dos modelos IA, visto que o número de algarismos após a vírgula é limitado. Foi realizada essa divisão dos dados de forma a obter uma melhor distribuição gráfica dos resíduos.

TABELA 2 – COEFICIENTES E INDICADORES ESTATÍSTICOS DOS MODELOS VOLUMÉTRICOS AJUSTADOS POR CLASSE DE DAP

Classe de dap (cm)	β_0	β_1	β_2	Conj.	R	RSME%
5 a 10	-9,729**	1,898**	0,941**	1	0,977	10,088
				2	0,988	7,645
> 10	-9,733**	1,853**	0,985**	1	0,988	7,827
				2	0,988	7,271

FONTE: A autora (2016).

Nota: Conj. 1 e 2 – Conjunto de treino (1) e validação (2); * significativo a 5%; ** significativo a 1%; ns não significativo.

Em complementaridade, os resíduos destes modelos ajustados foram avaliados quanto às condicionantes da regressão, sendo que o teste de White (W) avalia a homocedasticidade da variância, Shapiro-Wilk (SW) avalia a normalidade dos resíduos e Durbin-Watson (DW) avalia a independência destes, conforme a TABELA 3.

TABELA 3 – AVALIAÇÃO DAS CONDICIONANTES DA REGRESSÃO PARA OS RESÍDUOS DOS MODELOS DE SCHUMACHER E HALL AJUSTADOS POR CLASSE DE DAP

Classe de dap (cm)	W	SW	DW
5 a 10	4,03 ^{ns}	0,96 ^{ns}	1,924 ^{ns}
> 10	41,23 [*]	0,98 ^{ns}	1,768 ^{nc}

FONTE: A autora (2016).

Notas: * significativo a 5%; ns não significativo; nc não conclusivo; W – teste de White; SW – teste de Shapiro-Wilk; DW – teste de Durbin-Watson.

Conforme Gujarati (2006), premissas básicas referentes à utilização de modelos de regressão linear com estimativas dos parâmetros via mínimos quadrados são necessárias, dentre as quais destacam-se: linearidade dos parâmetros; resíduo com média zero; homocedasticidade da variância dos resíduos; ausência da autocorrelação dos termos de erro (independência dos resíduos); dados observados e erros não correlacionados; ausência de viés do modelo. A não observância dessas premissas denota que o modelo de regressão pode apresentar problemas em seus estimadores.

Assim, para o modelo referente aos indivíduos com dap superior a 10 cm é observada a significância no teste de White, rejeitando a hipótese de que os resíduos se distribuem de forma homogênea ao longo da linha estimativa, portanto, há heterocedasticidade. Como medida corretiva Gujarati (2006) sugere a utilização dos mínimos quadrados generalizados, que consistem na atribuição de pesos no ajuste dos parâmetros. Essa abordagem não será aplicada na presente pesquisa.

5.1.2 Estimativa do volume total obtido em função do dap e altura total

Os dois modelos de volume foram agrupados para facilitar o entendimento e visualização do seu desempenho, gerando estatísticas únicas para o modelo de regressão de Schumacher e Hall. Assim, a TABELA 4 apresenta as estatísticas de avaliação dos modelos, tanto para o grupo de teste utilizado no treino e ajuste, quanto para o grupo de validação.

TABELA 4 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO VOLUME TOTAL DO FUSTE

Conj.	Modelo	Estatísticas								Ranking
		R	RSME	RSME%	D	D	SD	SSRR	RP	
1	S & H	0,994	0,007	9,40	$-7 \cdot 10^{-5}$	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	2,22	5,65	4
	k-NN	0,999	0,002	2,18	$-4 \cdot 10^{-5}$	10^{-3}	$7 \cdot 10^{-5}$	0,24	1,82	1
	RF	0,998	0,005	6,05	10^{-4}	$2 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	1,21	4,09	2
	RNA	0,995	0,006	8,27	$-4 \cdot 10^{-4}$	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	2,49	6,54	3
2	S & H	0,995	0,007	8,60	$3 \cdot 10^{-4}$	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	1,10	5,34	2
	k-NN	0,993	0,008	9,84	$-1 \cdot 10^{-4}$	$5 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	1,59	6,31	3
	RF	0,990	0,009	11,74	$9 \cdot 10^{-4}$	$5 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	2,22	7,22	4
	RNA	0,995	0,006	8,22	$-2 \cdot 10^{-5}$	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	1,42	6,11	1

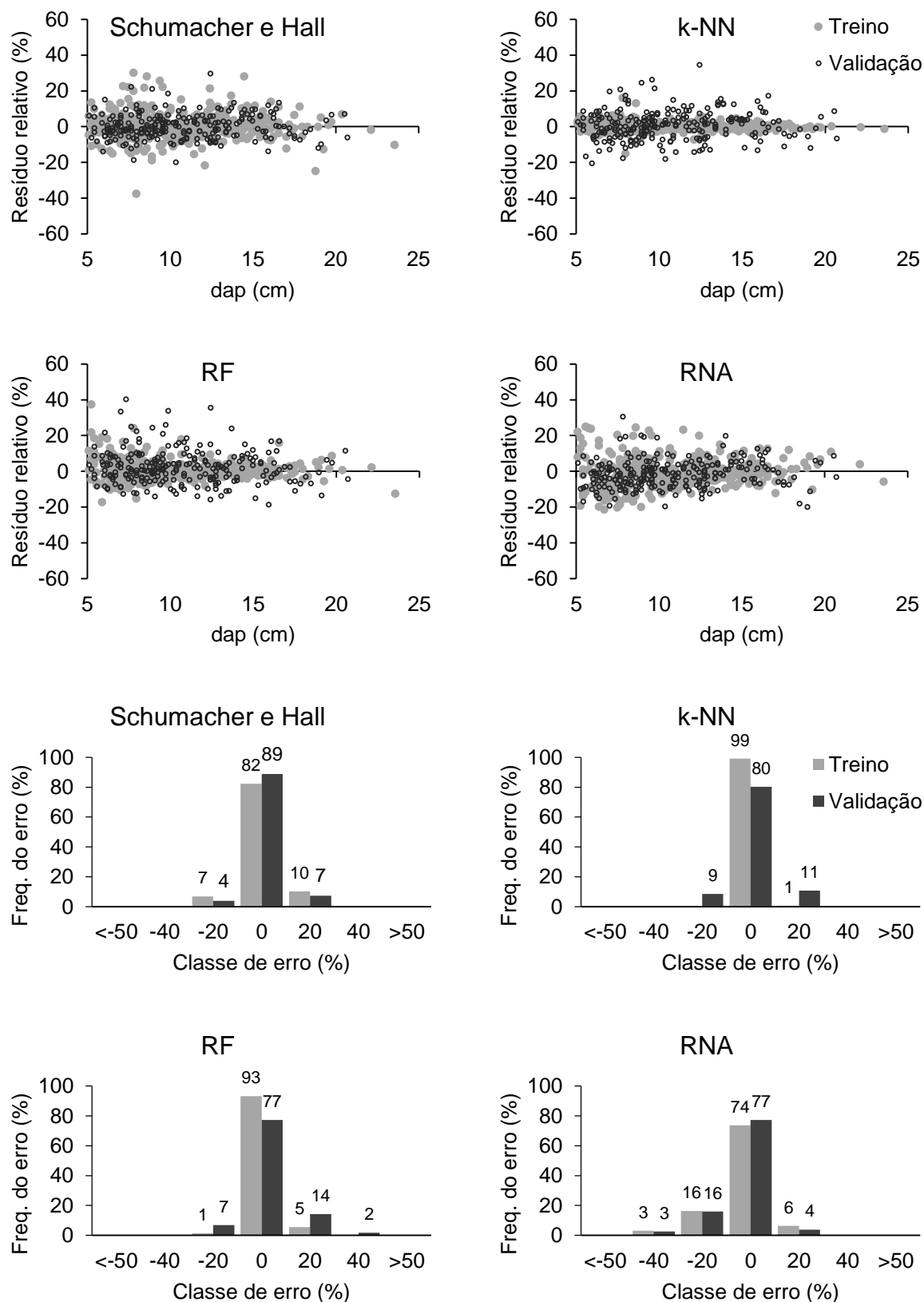
FONTE: A autora (2016).

Notas: Conj. 1 e 2 – Conjunto de treino (1) e validação (2); r – coeficiente de correlação entre valores observados e estimados; RSME e RSME% – raiz quadrada do erro médio absoluto em porcentagem; D – desvio médio; |D| – desvio médio absoluto; SD – desvio padrão das diferenças; SSRR – soma do quadrado do resíduo relativo; RP – porcentagem dos resíduos; Ranking com atribuição da menor nota ao modelo de melhor desempenho na estatística em análise.

Todos os modelos propiciaram estatísticas favoráveis à sua utilização, com coeficientes de correlação dos valores observados e estimados superiores a 0,99 e RSME baixos. Como esperado, as estatísticas do conjunto de teste demonstram maior precisão. Quanto ao ranqueamento dos modelos, para o conjunto de treino, os três modelos de aprendizado de máquina se mostraram mais acurados que Schumacher e Hall. Porém ao analisarmos o grupo de validação observa-se que esses modelos apresentam estatísticas mais instáveis, ou seja, Schumacher e Hall e as RNAs apresentam estatísticas muito similares nos dois grupos, enquanto os demais não. A exemplo do algoritmo vizinho mais próximo (k-NN), o qual foi eleito como o melhor modelo no conjunto de teste, porém, não obteve o mesmo êxito no conjunto de validação, ainda que suas estatísticas de erros apontem para um bom desempenho da técnica (alto r e baixo RSME).

Nesse sentido, o algoritmo k-NN apresenta a tendência de grande eficiência para o conjunto de teste e maior dispersão para a validação, conforme evidenciado nos resíduos plotados em relação ao seu valor observado (FIGURA 6), em que o viés apresentado no conjunto de teste é extremamente sutil e muito próximo a zero.

FIGURA 6 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO VOLUME TOTAL EM FUNÇÃO DO DAP E ALTURA TOTAL



FONTE: A autora (2016).

Por outro lado, tanto o modelo de Schumacher e Hall quanto a RNA (quando com treinamento adequado), têm por característica a formação de modelos, ainda que a RNA seja menos explícita e mais flexível devido ao alto número de neurônios, mas ainda assim ela tem como princípio a modelagem a fim de obter linhas médias para minimizar o viés e não ficar supertreinada para a base de dados.

A FIGURA 6 demonstra ainda que os modelos não apresentaram resíduos extremos, com suas respectivas frequências de erros entre as classes de centro $\pm 20\%$. Destaca-se aqui uma leve tendência da RNA em subestimar os volumes, especialmente os de menor magnitude.

A utilização de RNAs para estimativa do volume no Brasil foi introduzida por Gorgens et al. (2009) e Silva et al. (2009), ambos com abordagem metodológica similar, em que é proposta a utilização das redes como alternativa a modelos volumétricos consagrados, como Schumacher e Hall (1933). Os autores são unânimes que além da precisão similar ou mesmo superior ao modelo de regressão, as RNAs têm como vantagem capacidade de generalização e plasticidade que permite a utilização de apenas uma rede para realizar a predição de volume de árvores de distintos locais e clones diferentes, resultado esse de grande interesse ao modelador, não sendo, portanto, necessário o ajuste de diversos modelos volumétricos por clone, mas sim tendo o ajuste de RNA representativo para todos os indivíduos.

Gorgens et al. (2014) reportam resultados promissores com a utilização de RNAs treinadas com o algoritmo *backpropagation*, em que sugerem que a rede seja construída com mais de 10 neurônios na primeira camada, bem como recomendam o uso de mais de uma camada intermediária. Porém na presente pesquisa a otimização da rede não obteve um modelo complexo como o melhor resultado, visto que a rede tida como a de melhor desempenho possui 3 neurônios na camada intermediária. Resultado esse provavelmente devido à diferença nas bases de dados utilizadas, visto que no estudo de Gorgens et al. (2014) foram utilizados dados de distintos clones de *Eucalyptus* sp.

Özçelik et al. (2010) seguem a mesma linha de resultado quanto a simplicidade da RNA treinada via *backpropagation*, em que para a obtenção do volume obteve melhores resultados com redes simples, com uma camada intermediária e 2 ou 3 neurônios. Ao modelar quatro espécies florestais obteve RSME% entre 4,66 e 14,92%, mas relata melhores resultados com outro tipo de rede

neural, a correlação em cascada, destacando como vantagens o algoritmo ser mais rápido e com melhor convergência em relação ao *backpropagation*. Os autores destacam ainda como grande vantagem na utilização das RNAs, o fato destas conseguirem automaticamente assimilar relações entre variáveis através das conexões dos seus respectivos pesos, capacitando, portanto, as redes a modelar sistemas com relações não lineares complexas, sistemas esses que a análise de regressão convencional pode não ser o método mais acurado.

5.2IA x FUNÇÃO DE AFILAMENTO

5.2.1 Ajuste e teste das condicionantes de regressão da função de afilamento polinômio de Hradetzky

A TABELA 5 apresenta os coeficientes e suas respectivas potências selecionadas pelo método *stepwise*, sendo todos os coeficientes significativos ao menos a 95% de probabilidade. Os indicadores de ajuste como a correlação entre valores observados e preditos e RSME% demonstram que os modelos são adequados para a utilização na estimativa da forma do fuste. É observado uma maior dificuldade do modelo em expressar a forma em indivíduos na classe de dap menor que 5 cm.

TABELA 5 – COEFICIENTES, POTENCIAS SELECIONADAS PELO STEPWISE E INDICADORES ESTATÍSTICOS DAS FUNÇÕES DE AFILAMENTO AJUSTADAS POR CLASSE DE DAP

Classe de dap (cm)	β_0	β_1	β_2	β_3	β_4	Conj.	r	RSME%
< 5	30,560**	-29,421**	-0,717**	-0,402**	-	1	0,959	25,246
		0,005	1	4		2	0,950	25,964
5 - 10	18,096**	-17,256**	-0,734**	-0,168**	-	1	0,979	9,551
		0,005	2	25		2	0,982	8,992
> 10	18,382**	-17,598**	-0,613**	-0,374**	0,189*	1	0,988	7,565
		0,005	2	10	15	2	0,987	7,656

FONTE: A autora (2016).

Nota: Conj. 1 e 2 – Conjunto de treino (1) e validação (2); * significativo a 5%; ** significativo a 1%; ns não significativo.

Em complementaridade, os resíduos destes modelos ajustados foram avaliados quanto às condicionantes da regressão, sendo que o teste de White (W) avalia a homocedasticidade da variância, Shapiro-Wilk (SW) avalia a normalidade dos resíduos e Durbin-Watson (DW) avalia a independência destes, conforme TABELA 6. Conforme a estatística DW, os modelos apresentam tendenciosidade nos resíduos, sendo as demais estatísticas não significativas.

TABELA 6 – AVALIAÇÃO DAS CONDICIONANTES DA REGRESSÃO PARA OS RESÍDUOS DAS FUNÇÕES DE AFILAMENTO POR CLASSE DE DAP

Classe de dap (cm)	W	SW	DW
> 5	39,50 ^{ns}	0,75 ^{ns}	0,508*
5 a 10	65,05 ^{ns}	0,96 ^{ns}	0,662*
< 10	82,51 ^{ns}	0,96 ^{ns}	0,739*

FONTE: A autora (2016).

Notas: * significativo a 5%; ** significativo a 1%; ns não significativo; W – teste de White; SW – teste de Shapiro-Wilk; DW – teste de Durbin-Watson.

Conforme anteriormente mencionado, modelos polinomiais são eficientes na descrição do perfil total do fuste, porém podem resultar em distorções nas estimativas da base e da ponta (MAX; BURKHART, 1976; DEMAERSCHALK; KOZAK, 1977; JIANG et al., 2005; ROJO et al., 2005; BROOKS et al., 2008), em que, no presente trabalho, essa tendenciosidade é devida a dificuldade do modelo polinomial em estimar a porção final do fuste. Eisfeld et al. (2004) reiteram que o uso de modelos polinomiais é vantajoso devido a aliar eficiência a simplicidade de aplicação. Porém salientam também que estes apresentam algumas deficiências na representação da base, podendo comprometer as estimativas.

5.2.2 Estimativa do diâmetro ao longo do fuste

Os três modelos de afilamento foram agrupados para facilitar o entendimento e visualização do seu desempenho, gerando estatísticas únicas para o modelo de regressão de Hradetzky. Assim, a TABELA 7 apresenta as estatísticas de avaliação dos modelos, tanto para o grupo de teste utilizado no treino e ajuste, quanto para o grupo de validação.

TABELA 7 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO DIÂMETRO AO LONGO DO FUSTE

Conj.	Modelo	Estatísticas								Ranking
		r	RSME	RSME%	D	D	SD	SSRR	RP	
1	Hradetzky	0,990	0,58	9,25	0,04	0,40	0,01	159,68	10,77	4
	k-NN	0,999	0,11	1,78	$4 \cdot 10^{-5}$	0,07	10^{-3}	9,55	2,24	1
	RF	0,999	0,20	3,24	10^{-3}	0,14	$2 \cdot 10^{-3}$	35,99	4,21	2
	RNA	0,992	0,52	8,25	-0,06	0,36	0,01	170,75	9,76	3
2	Hradetzky	0,990	0,58	9,15	0,05	0,40	0,01	117,98	11,21	3
	k-NN	0,987	0,63	10,06	0,03	0,44	0,01	251,93	12,57	4
	RF	0,991	0,53	8,39	0,02	0,36	0,01	148,71	10,33	2
	RNA	0,992	0,51	8,04	-0,06	0,35	0,01	97,41	9,76	1

FONTE: A autora (2016).

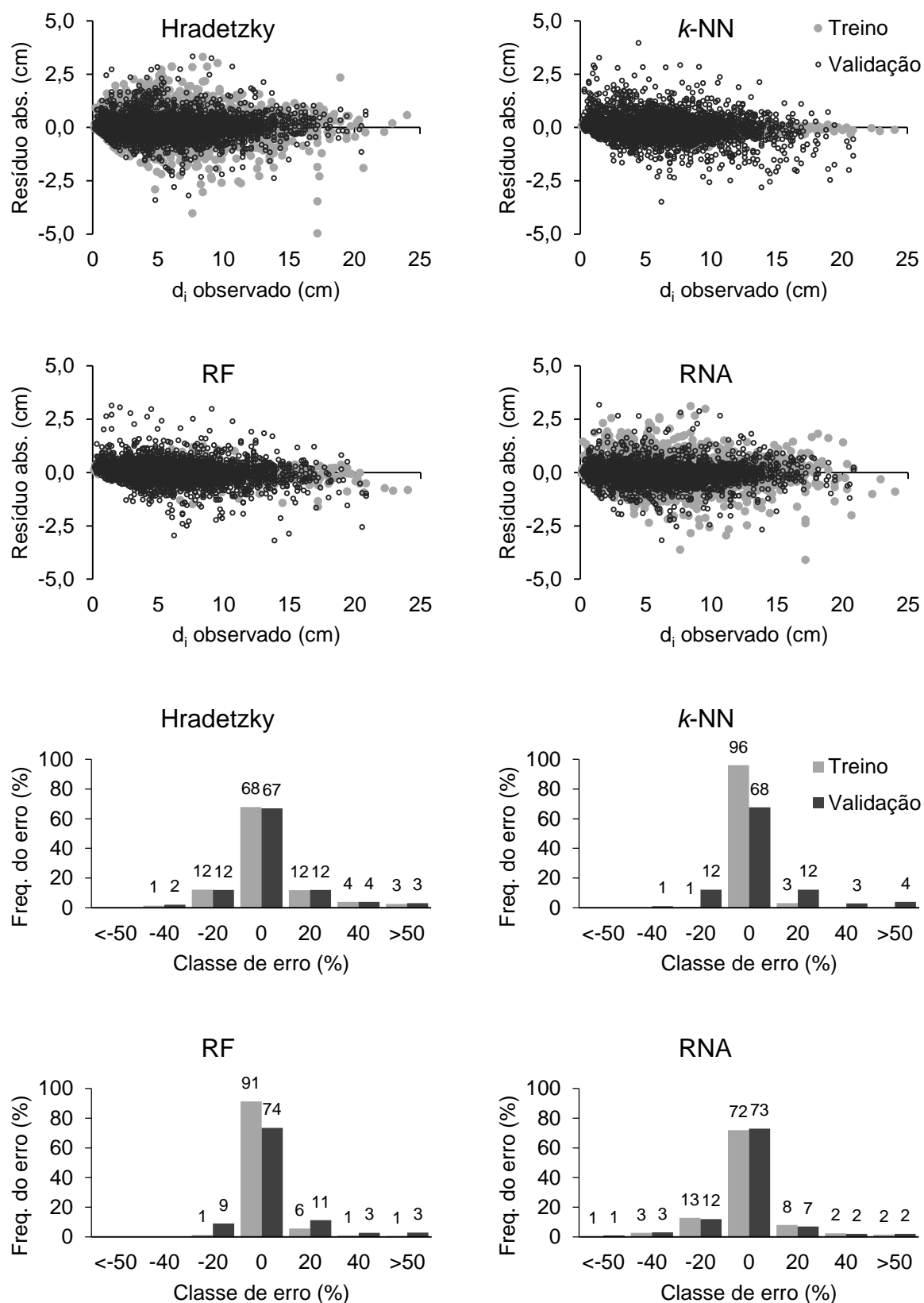
Notas: Conj. 1 e 2 – Conjunto de treino (1) e validação (2); r – coeficiente de correlação entre valores observados e estimados; RSME e RSME% – raiz quadrada do erro médio absoluto em porcentagem; D – desvio médio; |D| – desvio médio absoluto; SD – desvio padrão das diferenças; SSRR – soma do quadrado do resíduo relativo; RP – porcentagem dos resíduos; Ranking com atribuição da menor nota ao modelo de melhor desempenho na estatística em análise.

Todos os modelos propiciaram estatísticas favoráveis à sua utilização, com coeficientes de correlação superiores a 0,98 e RSME máximo de 10%. Como esperado, as estatísticas do conjunto de teste demonstram maior precisão. Cabe ressaltar que apenas essas duas estatísticas usualmente utilizadas não retratam adequadamente as funções de afilamento, pois tratam-se de médias que não levam em consideração a posição da estimativa ao longo do fuste, servindo, portanto, apenas como um indicativo da correlação entre variáveis (SILVA et al., 2011).

Quanto às estatísticas adicionais dos resíduos, todos os modelos apresentaram homogeneidade dos resíduos (SD muito próximo a zero), sem fortes tendências (D e |D|), e com resíduos de baixa amplitude (RP – similar ao RSME). A estatística Soma do quadrado do resíduo relativo (SSRR), por sua vez, representa o tamanho de cada resíduo em relação ao seu valor real, evidenciou maior discrepância entre os modelos devido a sua natureza quadrática.

A FIGURA 7 demonstra que os modelos não apresentaram resíduos extremos, com suas respectivas frequências de erros entre as classes de centro $\pm 20\%$.

FIGURA 7 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO DIÂMETRO AO LONGO DO FUSTE



FONTE: A autora (2016)

Quanto ao ranqueamento dos modelos, assim como na estimativa do volume em função do dap e altura, fica evidente que, ainda que tenha sido aplicada a validação cruzada no treinamento dos modelos de IA, estes podem não necessariamente incidir em bons ajustes no grupo de validação. Repetindo aqui a tendência do algoritmo vizinho mais próximo (k -NN), como o melhor modelo no conjunto de teste e pior modelo no conjunto de validação.

Leite et al. (2011) compararam o desempenho de três tipos de RNAs, tipo perceptron linear, perceptron múltiplas camadas e função de base radial, com a função de afilamento polinômio de segundo grau, consagrada como modelo de Kozak (1969). Destacam que os modelos de RNA apresentaram um melhor ajuste a base de dados, com baixos valores de RSME% e concentração dos resíduos entre $\pm 10\%$. Porém, assim como a presente pesquisa, os modelos apresentaram dificuldades nas estimativas dos menores diâmetros situados na porção terminal do fuste, com fortes tendências a superestimativas.

Soares et al. (2011) reportam bom desempenho das RNAs na estimativa de diâmetros relativos treinados com a mesma metodologia deste trabalho, com coeficientes de correlação entre 0,97 e 0,99, bem como valores de RSME% de 7% em média. Como diferencial ainda apontam boa acurácia das RNAs na predição de diâmetros recursivos, ou seja, com a mensuração de apenas 3 medidas da base na etapa de generalização (validação) do modelo e posterior estimativa dos demais. Assim, seria possível a obtenção de valores mais acurados dos sortimentos obtidos em inventários florestais, que comumente são tomadas apenas as medidas do dap e altura total. Soares et al. (2012) ainda propõem a utilização de RNAs para estudo da forma sem o conhecimento prévio da altura total, com coeficientes de correlação entre 0,95 e 0,99, bem como RSME% médios entre 1% e 20%, considerando uma estratificação dos indivíduos por classe de dap.

Özçelik et al. (2014) comparou o desempenho de três tipos de RNAs, dentre elas perceptrons múltiplas camadas treinadas por *backpropagation*, em relação a um modelo de afilamento não linear na estimativa dos diâmetros relativos e volume. Esses autores destacam como vantagem que as RNAs conseguem assimilar as relações existentes entre as variáveis de entrada e de saída automaticamente através das conexões dos pesos, viabilizando assim a modelagem de interações não lineares complexas, ao passo que modelos de regressão não possuem tal poder

preditivo. Assim, as RNAs vêm ganhando notoriedade em aplicações matemáticas no inventário florestal justamente por essa plasticidade na modelagem.

Conforme já anteriormente mencionado, as aplicações dos algoritmos *k*-vizinho mais próximo e *Random Forest* como aproximação de função/regressão são escassas, portanto não foi possível fazer um cenário comparativo para a estimativa do diâmetro com esses algoritmos.

5.2.3 Estimativa do volume total

A TABELA 8 apresenta as estatísticas de avaliação dos modelos, tanto para o grupo de teste utilizado no treino e ajuste, quanto para o grupo de validação. Todos os modelos apresentaram coeficientes de correlação superiores a 0,98, porém com valores de RSME% superiores em relação aos obtidos nas estimativas do diâmetro relativo, indicando maior variabilidade nas estimativas. A FIGURA 8 apresenta os gráficos de resíduos para o volume total, evidenciando que o modelo *k*-NN apresentou um maior viés tanto para o conjunto de treino quanto para o conjunto de validação. Juntamente com o polinômio de Hradetzky, é possível evidenciar que ambos apresentam leve tendência a superestimar os volumes, sendo que o *k*-vizinho mais próximo apresenta um viés mais forte.

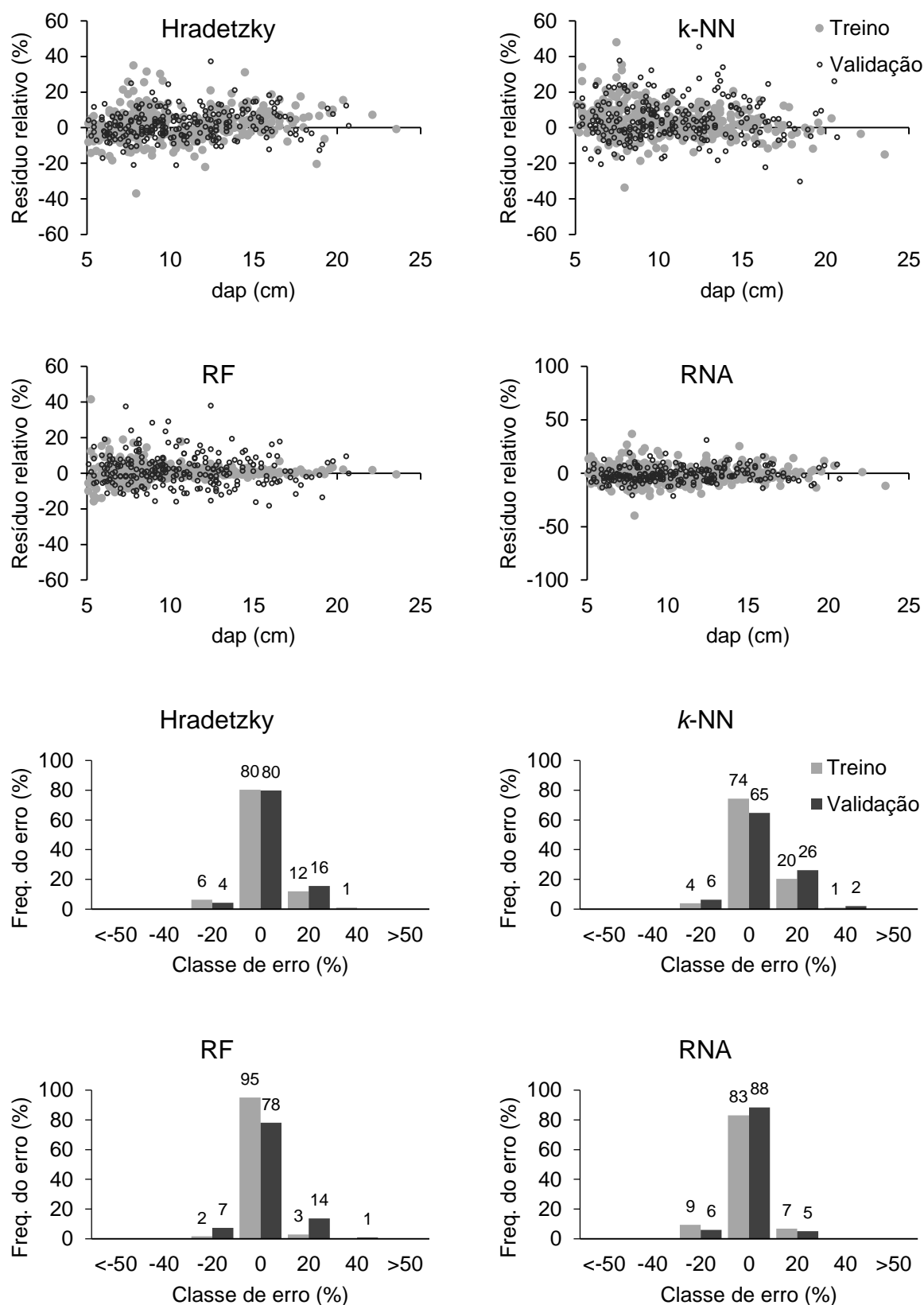
TABELA 8 – ESTATÍSTICAS DE AJUSTE PARA A ESTIMATIVA DO VOLUME TOTAL DO FUSTE

Conj.	Modelo	Estatísticas								Ranking
		r	RSME	RSME%	D	D	SD	SSRR	RP	
1	Hradetzky	0,994	0,01	10,40	$2 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	2,67	6,35	4
	<i>k</i> -NN	0,994	0,01	10,05	10^{-3}	$5 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	3,29	7,20	3
	RF	0,999	$2 \cdot 10^{-3}$	2,17	$3 \cdot 10^{-4}$	10^{-3}	$6 \cdot 10^{-5}$	0,85	2,97	1
	RNA	0,994	0,01	9,31	10^{-3}	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	2,30	5,95	2
2	Hradetzky	0,994	0,01	10,80	$2 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	1,55	6,24	2
	<i>k</i> -NN	0,989	0,01	12,64	$2 \cdot 10^{-3}$	0,01	$5 \cdot 10^{-4}$	3,43	8,97	4
	RF	0,990	0,01	11,65	$8 \cdot 10^{-4}$	0,01	$5 \cdot 10^{-4}$	1,94	6,78	3
	RNA	0,995	0,01	8,32	$6 \cdot 10^{-4}$	$4 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	1,13	5,47	1

FONTE: A autora (2016).

Notas: Conj. 1 e 2 – Conjunto de treino (1) e validação (2); r – coeficiente de correlação entre valores observados e estimados; RSME e RSME% – raiz quadrada do erro médio absoluto em porcentagem; D – desvio médio; |D| – desvio médio absoluto; SD – desvio padrão das diferenças; SSRR – soma do quadrado do resíduo relativo; RP – porcentagem dos resíduos; Ranking com atribuição da menor nota ao modelo de melhor desempenho na estatística em análise.

FIGURA 8 – DISPERSÃO DOS RESÍDUOS RELATIVOS E FREQUÊNCIA DOS RESÍDUOS PORCENTUAIS POR CLASSE DE ERRO PARA ESTIMATIVA DO VOLUME TOTAL DO FUSTE



FONTE: A autora (2016)

Quanto ao ranqueamento, o modelo RF se mostrou mais acurado para o conjunto de treino, seguido da RNA. No conjunto de validação, por sua vez, a RNA demonstrou ser o modelo de melhor desempenho, seguido do polinômio Hradetzky. O modelo k -NN não propiciou bom resultado no ranqueamento como na estimativa do diâmetro relativo.

Os modelos RF e RNA apresentaram uma distribuição mais homogênea e com seus pontos próximos a zero, bem como pode ser visualizado na distribuição da frequência dos resíduos que os modelos RF e RNA apresentaram maior porção dos resíduos na classe que varia entre $\pm 10\%$, bem como não apresentaram tendência a super ou subestimar.

Özçelik et al. (2010) indicam que a implementação de RNAs na estimação florestal oferece diversas vantagens em relação a métodos tradicionalmente abordados, destacando que problemas como supertreinamento podem ser facilmente evitados com a seleção de arquiteturas adequadas e com a utilização de base de dados de ajuste e treino. Esse fato pode ser evidenciado na presente pesquisa, em que a RNA se mostrou como o melhor modelo dentre os analisados.

Binoti et al. (2014b) testaram um método para obtenção de estimativas de volume total com e sem casca para clones de *Eucalyptus* sp. utilizando RNAs. Para tanto, as redes foram treinadas com as variáveis clone, dap, altura total e diâmetros relativos às alturas 0,5, 1, 2 e 4 m. Esses autores apontam que é possível obter estimativas muito acuradas utilizando as RNAs, com valores de r acima de 0,99, RSME inferiores a 5% e resíduos situados entre $\pm 10\%$. Assim, afirmam que essa metodologia pode ser aplicada como alternativa para redução das medições de diâmetro ao longo do fuste, visando a cubagem das árvores em pé, mas ainda assim mantendo os níveis de precisão aceitáveis.

Sanquetta et al. (2015b) testaram o desempenho de modelos k -NN para a estimativa de volume de colmos de *Bambusa* sp., juntamente com as funções de afilamento polinômio de quinto grau e potências fracionárias de Hradetzky, além de outras técnicas consagradas para a estimativa de volume do fuste. Esses autores consideram que não foram obtidos resultados satisfatórios com os modelos k -NN, denominados no trabalho de método de mineração de dados, os quais não apresentaram resíduos com distribuição homogênea, mas sim com forte heterocedasticidade e grande dispersão. Assim, atribuem o mau desempenho na estimativa ao número reduzido da base de dados, destacando ainda que o uso da

técnica de mineração de dados nesse caso só seria indicado caso as premissas da regressão linear fossem violadas.

5.3 ABORDAGEM COMPARATIVA DOS MÉTODOS ANALISADOS

Visto que a presente pesquisa introduz a utilização de métodos de aprendizado de máquina como uma alternativa ao método consagrado na literatura no estudo da forma do fuste e seu volume, o QUADRO 1 apresenta um resumo comparativo das quatro técnicas de modelagem utilizadas na presente pesquisa, visando destacar assim suas potencialidades e fragilidades.

De maneira geral, conforme Strobl et al. (2009), os métodos não paramétricos, tais como os de IA utilizados no presente estudo, possuem a habilidade de identificar relações entre variáveis que são muito complexas para os modelos paramétricos, assim como, possuem mais facilidade em trabalhar com um grande número de variáveis explicativas.

Haara e Kangas (2012) afirmam que conforme os dados utilizados sejam desequilibrados e com forte não linearidade, o k -NN apresenta resultados menos tendenciosos que a regressão convencional. Por outro lado, Maltamo et al. (1998) ao utilizarem o k -NN já atentaram que não é assegurado que as estimativas obtidas pelo algoritmo sejam livres de viés, bem como exigem uma base de dados muito mais extensa. Mognon et al. (2014) corroboram essa afirmação de que o tamanho da amostra pode ser um limitador à obtenção de estimativas acuradas pelo k -NN, indicando que quando não há disponibilidade de uma grande e robusta base de dados, a regressão deve ser preferida.

QUADRO 1 - COMPARAÇÃO DAS QUATRO TÉCNICAS DE MODELAGEM APLICADAS

Modelo	Método	Vantagem	Desvantagem
Análise de regressão (Modelos de Schumacher e Hall e Hradetzky)	Regressão linear pelo método dos mínimos quadrados, com seleção de variáveis significativas via <i>stepwise</i> .	Teoria sólida e com vasta aplicação. Possibilidade de obtenção de distintas variáveis de interesse (di, h, v). Função integrável.	Comportamento da curva previamente definido, dificultando assim o ajuste adequado para distintas bases de dados.
<i>k</i> -NN	Métricas de distâncias tratando observação a observação.	Grande plasticidade na modelagem de relações complexas.	Necessário grande base homogênea de dados, do contrário pode apresentar resultados tendenciosos.
RF	Construção de árvores de regressão independentes entre si, que trazem por si só grande poder preditivo.	Possui alto poder preditivo, bem como trabalha facilmente com variáveis categóricas.	Esse alto poder preditivo ainda é pouco compreendido pelos trabalhos já publicados.
RNA	Diversos algoritmos de treinamento, com ajuste dos pesos que se conectam à rede conforme as variações na taxa de erro.	Capaz de modelar relações complexas, lidar com variáveis categóricas e/ou grande quantidade de variáveis explicativas, sejam de grandezas similares ou não. Rapidez na aplicação (fase posterior ao treino).	Exigência computacional para otimização de seus parâmetros na fase de treino. Demanda cuidados a fim de evitar o supertreinamento a base de dados. Dificuldade na escolha do algoritmo de treinamento.

FONTE: A autora (2016).

Fehrmann et al. (2008), ao compararem estimativas de biomassa obtidas pelo *k*-NN e modelagem linear mista, traçaram um interessante paralelo acerca dos dois métodos: destacam a habilidade do método vizinho mais próximo em explorar desvios locais na base de dados, capacidade essa que a regressão não possui. Porém destacam ainda que a qualidade das estimativas é altamente dependente do número de observações e sua distribuição. Esses autores ressaltam que a vantagem da regressão é óbvia: uma vez que os parâmetros foram estimados, estes são facilmente aplicados, mas destacam também que no método paramétrico a generalização é mais limitada, ao passo que o método não paramétrico *k*-NN pode ser mais facilmente “calibrado”.

Ao que concerne a utilização das RNAs, conforme já foi anteriormente abordado, sua utilização está mais difundida na modelagem florestal que os demais algoritmos aqui testados, porém há muito ainda a explorar.

Em estudos para obtenção de volumes do fuste com a utilização de RNAs, trabalhos como Gorgens et al. (2009), Silva et al. (2009) e Binoti et al. (2014a, b) são unânimes em apontar como principais vantagens das redes a habilidade em modelar características não explícitas entre as variáveis, além da capacidade de com um modelo único obter estimativas para diferentes estratos, ao passo que modelos de regressão exigem um modelo para cada estrato, onerando o processamento e análise. Binoti et al. (2014a) destacam também como importante vantagem a facilidade na inclusão de variáveis categóricas nos modelos de RNA.

Leite et al. (2011) ressaltam como principais vantagens que as RNAs possuem grande adaptabilidade, tolerância a ruídos e *outliers*, grande habilidade no aprendizado e consequentemente bons resultados na generalização para dados desconhecidos. Porém, os autores apresentam também algumas limitações, tais como as dificuldades na escolha dos parâmetros de configuração da rede, como número de neurônios na camada oculta e a não garantia que o algoritmo apresentou a solução ótima para o modelo.

Assim como as RNAs, modelos RF possuem grande potencialidade de utilização na área florestal. Conforme Moreno-Fernández et al. (2015) destacam, dados de origem ecológica raramente satisfazem as premissas básicas para regressão, assim, os métodos não paramétricos apresentam grande potencial para utilização, possibilitando a modelagem dessa estrutura complexa. Porém os autores destacam como limitação a impossibilidade de interpretação de parâmetros como em um modelo tradicional.

Cutler et al. (2007) vêm pela mesma linha de pensamento, em que destacam que modelos com tamanha capacidade computacional não apresentam relações diretas entre variáveis preditoras e variáveis resposta, dificultando o entendimento dos processos dos algoritmos até o resultado final. Nesse tipo de modelo não é possível fazer a interpretação dos parâmetros, bem como há dependência direta do *software* utilizado, pois não há um modelo a ser obtido e facilmente replicado assim como modelos de regressão.

Prasad et al. (2006) apresentam como maior vantagem na utilização do RF que, dado o grande número de árvores utilizadas no modelo, este não apresenta

supertreinamento à base de dados. Destacam que a seleção do preditor aleatório mantém os desvios baixos e com melhores estimativas do modelo quando comparado a outros modelos que simulam árvores de regressão. Porém os autores também ressaltam como desvantagem o fato da alta capacidade preditiva do RF não é facilmente compreendida, bem como a forte exigência computacional (diretamente ligada ao tamanho e complexidade da base de dados).

De posse de conhecimento acerca das potencialidades e limitações dos métodos de obtenção de estimativas, é possível a tomada de decisão sobre qual o mais indicado a cada situação. Como já foi abordado, métodos de IA carecem ainda de estudos para solidificar sua utilização na área florestal, bem como exigem uma base de dados mais extensa e representativa. Porém, é de se destacar a potencialidade desses métodos quando as premissas da análise de regressão são burladas.

Por outro lado, a regressão convencional pode ser menos exigente em termos de número de observações no ajuste, porém pode ser mais limitada quando na obtenção dos resultados. Essa abordagem apresenta modelos mais regionalizados, ou um conjunto de diversas equações para situações diferentes, como os modelos por classe diamétrica utilizados na presente pesquisa. Nesse sentido os modelos de IA podem facilmente englobar diversas variações em um mesmo modelo, sem a necessidade da segmentação que muitas vezes a regressão exige.

6 CONCLUSÕES

Para a estimativa do volume em função do dap e altura, a RNA e o modelo de Schumacher e Hall apresentam melhores resultados no ranqueamento do que o *k*-vizinho mais próximo e *Random Forest*.

Os métodos aplicados de inteligência artificial demonstram ser mais acurados que o polinômio de Hradetzky para estimativas acerca do estudo da forma da árvore, tais como o diâmetro ao longo do fuste e volume total.

Os modelos de AM são adequados como alternativa na modelagem tradicionalmente aplicada na mensuração florestal, contudo a sua utilização deve ser cuidadosa devida a maior possibilidade de supertreinamento a base de ajuste.

7 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

O presente trabalho teve como objetivo uma comparação inicial e macro entre algoritmos de Aprendizado de Máquina e a técnica usual de regressão. Significando, portanto, que os algoritmos de AM não foram plenamente explorados em suas variantes, desse modo, para novas perspectivas de pesquisa, pode-se sugerir:

- Para o algoritmo k -NN, recomenda-se a análise de diferentes métricas de distância, estudo da sensibilidade na variação no número de vizinhos (k), diferentes métodos de ponderação dos k vizinhos selecionados;
- Para o algoritmo RF, recomenda-se uma análise da influência no número de árvores cultivadas;
- Para os modelos de RNA, é possível fazer uma análise ainda mais robusta, testando-se os diferentes tipos de RNAs, tais como a perceptron com uma camada, perceptron multiplas camadas (utilizada no presente trabalho), redes de base radial, redes auto-organizáveis, redes de aprendizagem extrema dentre mais uma infinidade de tipos de RNAs que podem ser analisadas. Das redes do tipo MLP, podem ser analisados diferentes algoritmos de treinamento além do *backpropagation*, tais como seus variantes diretos como *Resilient Propagation* RPROP+, dentre outras possibilidades de algoritmos. Somam-se as já citadas possibilidades, a variação do número de neurônios em uma ou mais camadas ocultas, diferentes funções de ativação e taxas de aprendizado.
- Além dos algoritmos aqui testados, pode-se destacar outros poderosos preditores que podem ser promissores à obtenção de estimativas da área florestal, com destaque a máquina de vetor suporte (SVM), lógica neuro-fuzzi, Redes bayesianas, algoritmos genéticos dentre muitos outros.

Sendo assim, o estudo da aplicação dos algoritmos de AM na área florestal está apenas sendo iniciado, em que análises quanto suas aplicações devem ser feitas para as mais diversas demandas de estimativas da área, possibilitando assim maior segurança na sua utilização.

REFERÊNCIAS

- AHA, D. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. **International Journal of Man-Machine Studies**, v. 36, n. 2, p. 267-287, 1992. [http://dx.doi.org/10.1016/0020-7373\(92\)90018-G](http://dx.doi.org/10.1016/0020-7373(92)90018-G)
- AHRENS, S.; HOLBERT, D. Uma função para forma de tronco e volume de *Pinus taeda* L. **Boletim de Pesquisa Florestal**, Colombo, n. 3, p. 37-68, 1981.
- ARAUJO, B. H. P. **Modelagem da altura, volume e afilamento do fuste de *Calycophyllum spruceanum* Benth. empregando regressão e redes neurais artificiais**. 53 f. Dissertação (Mestre em Ciências de Florestas Tropicais) – Instituto Nacional de Pesquisas da Amazônia, 2015.
- ARCE, J. E.; KOEHLER, A.; JASTER, C. B.; SANQUETTA, C. R. **Florexel – Funções Florestais desenvolvidas para o Microsoft Excel**. Centro de Ciências Florestais e da Madeira – CCFM – Universidade Federal do Paraná. Curitiba, 2000.
- ATTOH-OKINE, N. O. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. **Advances in Engineering Software**, v. 30, n. 4, p. 291-302, 1999. [http://dx.doi.org/10.1016/S0965-9978\(98\)00071-4](http://dx.doi.org/10.1016/S0965-9978(98)00071-4)
- BAFFETTA, F.; FATTORINI, L.; FRANCESCHI, S.; CORONA, P. Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. **Remote Sensing of Environment**, v. 113, n. 3, p. 463-475, 2009. <http://dx.doi.org/10.1016/j.rse.2008.06.014>
- BEIJMA, S.; COMBER, A.; LAMB, A. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. **Remote Sensing of Environment**, v. 149, p. 118-129, 2014. <http://dx.doi.org/10.1016/j.rse.2014.04.010>
- BLANCHETTE, D.; FOURNIER, R. A.; LUTHER, J. E.; CÔTÉ, J. F. Predicting wood fiber attributes using local-scale metrics from terrestrial LiDAR data: A case study of Newfoundland conifer species. **Forest Ecology and Management**, v. 347, p. 116-129, 2015. <http://dx.doi.org/10.1016/j.foreco.2015.03.013>
- BINOTI, M. L. M. S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 639-645, 2013. <http://dx.doi.org/10.1590/S0100-67622013000400007>
- BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira**, Pelotas, v. 5, n. 1, p. 58-67, 2014a. <http://dx.doi.org/10.12953/2177-6830.v05n01a06>
- BINOTI, M. L. M. S.; BINOTI, D. H. B.; LEITE, H. G.; GARCIA, S. L. R.; FERREIRA, M. Z.; RODE, R.; SILVA, A. A. L. Redes neurais artificiais para estimação do volume

de árvores. **Revista Árvore**, Viçosa, v. 38, n. 2, p. 283-288, 2014b.
<http://dx.doi.org/10.1590/S0100-67622014000200008>

BLANCO, A. M.; SOTTO, A.; CASTELLANOS, A. Prediction of the Amount of Wood Using Neural Networks. **Journal of Mathematical Modelling and Algorithms**, v. 11, n. 3, p. 295-307, 2012

BRADZIL, P.; SOARES, C. C.; JOAQUIM, P. Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results. **Machine Learning**, v. 50, p. 251-277, 2003. <http://dx.doi.org/10.1023/A:1021713901879>

BRANDÃO, F. G. **Estimativa da altura total de *Eucalyptus* sp. utilizando lógica fuzzy e neuro fuzzy**. Monografia (Bacharel em Ciência da Computação) – Universidade Federal de Lavras, 2007.

BRASIL. Ministério da Agricultura e Reforma Agrária. Secretaria Nacional de Irrigação. Departamento Nacional de Meteorologia. **Normais climatológicas (1961-1990)**. Brasília, 1992. 84p.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC, 2007. 260p.

BREIDENBACH, J.; NOTHDURFT, A.; KANDLER, G. Comparison of nearest neighbour approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. **European Journal of Forest Research**, v. 129, n. 5, p. 833-846, 2010.
<http://dx.doi.org/10.1007/s10342-010-0384-1>

BREIMAN, L. Bagging predictors. **Machine learning**, v. 24, n. 2, p. 123-140, 1996.
<http://dx.doi.org/10.1023/A:1018054314350>

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.
<http://dx.doi.org/10.1023/A:1010933404324>

BROOKS, J. R.; JIANG, L.; OZÇELIK, R. Compatible stem volume and taper equations for Brutian pine, Cedar of Lebanon, and Cilicica fir in Turkey. **Forest Ecology and Management**, Amsterdam, v. 256, p. 147-151, 2008.
<http://dx.doi.org/10.1016/j.foreco.2008.04.018>

CAI, S.; KANG, X.; ZHANG, L.; GONG, Z.; QIN, L.; CHEN, P. A Model for Tree Diameter Distribution in Stands Based on Artificial Neural Network. In: 2010 International Symposium on Intelligence Information Processing and Trusted Computing. Disponível em <
<http://ieeexplore.ieee.org/xpl/abstractKeywords.jsp?arnumber=5663228>> Acesso em 24/10/2015.

CAMPOS, J. C. C.; LEITE, H. G.; **Mensuração Florestal: Perguntas e Respostas**. 2.ed. – Viçosa: ed. UFV, 2006. 470p.

CARVALHO, P. E. R. **Espécies arbóreas brasileiras**. Brasília, DF: Embrapa Informação Tecnológica; Colombo, PR: Embrapa Florestas, 2003.

CASTAÑO-SANTAMARÍA, J.; CRECENTE-CAMPO, F.; FERNÁNDEZ-MARTÍNEZ, J. L.; BARRIO-ANTA, M.; OBESO, J. R. Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. **Forest Ecology and Management**, v. 307, p. 63-73, 2013. <http://dx.doi.org/10.1016/j.foreco.2013.07.014>

CHEN, W.; LI, X.; WANG, Y.; CHEN, G.; LIU, S. Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. **Remote Sensing of Environment**, v. 152, p. 291-301, 2014. <http://dx.doi.org/10.1016/j.rse.2014.07.004>

CHIRICI, G.; BARBATI, A.; CORONA, P.; MARCHETTI, M.; TRAVAGLINI, D.; MASELLI, F.; BERTINI, R. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. **Remote Sensing of Environment**, v. 112, n. 5, p. 2686-2700, 2008. <http://dx.doi.org/10.1016/j.rse.2008.01.002>

CORDEIRO, M. A.; PEREIRA, N. N. J.; BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, Colombo, v. 35, n. 83, p. 255-261, 2015.

CUTLER, D. R.; EDWARDS, T. C.; BEARD, K. H.; CUTLER, A.; HESS, K.; GIBSON, J.; LAWLER, J. Random forests for classification in ecology. **Ecology**, v. 88, p. 2783-2792, 2007. <http://dx.doi.org/10.1890/07-0539.1>

COVER, T. M.; HART, P. E. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n.1, p. 21–27, 1967.

CUNHA, N. G.; SILVEIRA, R. J. C.; SEVERO, C. R. S.; PINTO, L. F. S.; MENDES, R. G.; SILVA, J. B.; DUARTE, L. R.; SCHUMACHER, R. L. **Estudo de solos do município de Encruzilhada do Sul - RS**. Pelotas: EMBRAPA Clima Temperado, 2005. 82 p. (EMBRAPA Clima Temperado. Circular Técnica, 45)

CYBENKO, G. V. Approximation by Superpositions of a Sigmoidal function. **Mathematics of Control, Signals and Systems**, v. 2, n. 4, p.303-314, 1989. <http://dx.doi.org/10.1007/BF02134016>

DAVID, H. C.; MARINHESKI FILHO, A.; PELISSARI, A. L.; PÉLLICO NETTO, S.; ARAÚJO, E. J. G.; BAUM, L. Critérios de estratificação para o ajuste de funções de afilamento em fustes de pinus. **Pesquisa Florestal Brasileira**, Colombo, v. 34, n. 79, p. 197-206, 2014. <http://dx.doi.org/10.4336/2014.pfb.34.79.659>

DEMAERSCHALK, J. P.; KOZAK, A. The whole-bole system: a conditioned dual-equation system for precise prediction of tree profiles. **Canadian Journal of Forest Research**, Ottawa, v. 7, p. 488-497, 1977. <http://dx.doi.org/10.1139/x77-063>

DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. **Computers and Electronics in Agriculture**, v. 48, n. 3, p. 235-244. 2005. <http://dx.doi.org/10.1016/j.compag.2005.04.002>

DIAMANTOPOULOU, M. J.; MILIOS, E. Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. **Biosystems Engineering**, v. 105, n. 3, p. 306-315, 2010. <http://dx.doi.org/10.1016/j.biosystemseng.2009.11.010>

DIAMANTOPOULOU, M. J.; ÖZÇELİK, R. Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey. **Forest Systems**, v. 21, n. 3, p. 383-397, 2012. <http://dx.doi.org/10.5424/fs/2012213-02338>

DIAMANTOPOULOU, M. J.; ÖZÇELİK, R.; CRECENTE-CAMPO, F.; ELER, U. Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. **Biosystems Engineering**, v. 133, p. 33-45, 2015. <http://dx.doi.org/10.1016/j.biosystemseng.2015.02.013>

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2. ed. New York: John Wiley & Sons, 2001. 738p.

EISFELD, R. L.; MELLO, A. A.; SANQUETTA, C. R.; WEBER, K. S. Avaliação de modelos polinomiais na estimativa de volume total e por sortimento de *Pinus taeda*. **Brasil Florestal**, n. 79, p. 9-15, 2004.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 2.ed. Rio de Janeiro: EMBRAPA-SPI, 2006.

ESKELSON, B. N. I.; TEMESGEN, H.; LEMAY, V.; BARRETT, T. M.; CROOKSTON, N. L.; HUDAK, T. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. **Scandinavian Journal of Forest Research**, v. 24, p. 235-246, 2009. <http://dx.doi.org/10.1080/02827580902870490>

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011. 378p.

FEHRMANN, L.; LEHTONEN, A.; KLEINN, C.; TOMPPA, E. Comparison of linear and mixed-effect regression models and a k-nearest neighbour approach for estimation of single-tree biomass. **Canadian Journal of Forest Research**, v. 38, n. 1, p. 1-9, 2008. <http://dx.doi.org/10.1139/X07-119>

FIGUEIREDO, E. O. **Avaliação de povoamentos de teca (*Tectona grandis* L.f.) na microrregião do Baixo Rio Acre**. 301 f. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Lavras, Lavras, 2005.

FIGUEIREDO FILHO, A. **Influência da Resinagem no Crescimento de *Pinus elliottii* Engelm. var. *elliottii* e sua Avaliação Econômica**. 138p. Tese (Doutorado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 1991.

FIGUEIREDO FILHO, A.; BORDERS, B. E.; HITCH, K. L. Taper equations for *Pinus taeda* in southern Brazil. **Forest Ecology and Management**, Amsterdam, v. 83, n. 1-2, p. 39-46, 1996. [http://dx.doi.org/10.1016/0378-1127\(96\)03706-1](http://dx.doi.org/10.1016/0378-1127(96)03706-1)

FINGER, C. A. G. **Fundamentos de biometria florestal**. Santa Maria: UFSM, CEPEF-FATEC, 1992.

FIX, E.; HODGES, J. L. **Discriminatory analysis** – Nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.

FRIEDL, R. A. **Dinâmica da forma dos fustes em povoamentos plantados de *Araucaria angustifolia* (Bert) Ktze**. 166p. Dissertação (Mestrado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 1989.

FUCHS, H.; MAGDON, P.; KLEINN, C.; FLESSA, H. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. **Remote Sensing of Environment**, v. 113, n. 3, p. 518-531, 2009. <http://dx.doi.org/10.1016/j.rse.2008.07.017>

GORGENS, E. B.; LEITE, H. G.; GLERIANI, J. M.; SOARES, C. P. B.; CEOLIN, A. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, Viçosa, v. 33, n. 6, p. 1141-1147, 2009. <http://dx.doi.org/10.1590/S0100-67622009000600016>

GORGENS, E. B.; MONTAGHI, A.; RODRIGUEZ, L. C. E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers and Electronics in Agriculture**, v. 116, p. 221-227, 2015. <http://dx.doi.org/10.1016/j.compag.2015.07.004>

GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEEER, H. Soil organic carbon concentrations and stocks on Barro Colorado Island – Digital soil mapping using Random Forests analysis. **Geoderma**, v. 146, n. 1-2, p. 102-113, 2008. <http://dx.doi.org/10.1016/j.geoderma.2008.05.008>

GRINAND, C.; RAKOTOMALALA, F.; GOND, V.; VAUDRY, R.; BERNOUX, M.; VIEILLEDENT, G. Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. **Remote Sensing of Environment**, v. 139, p. 68-80, 2013. <http://dx.doi.org/10.1016/j.rse.2013.07.008>

GUJARATI, D. **Econometria básica**. São Paulo: Elsevier; 2006.

HAARA, A.; KANGAS, A. Comparing k nearest neighbours methods and linear regression – is there reason to select one over the other? **Mathematical and Computational Forestry & Natural-Resource Sciences**, v. 4, n. 1, p. 50-65, 2012. Disponível em < <http://mcfns.com/index.php/Journal/article/view/MCFNS.4%3A50>> Acesso em 10 de agosto de 2015.

HRADETZKY, J. **Analyse und interpretation statistischer abhänger keiten**. (Biometrische Beiträge zu aktueller forschungs projekten). Baden: Württemberg Mitteilungen der FVA, 1976. 146 p.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, 2009.

HAYKIN, S. S. **Redes neurais**: princípios e prática. Tradução Paulo Martins Engel. 2.ed. Porto Alegre: Bookman, 2001. 900p.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward network are universal approximators. **Neural Networks**, Bandera, v. 2, n. 5, p. 359-366, 1989. [http://dx.doi.org/10.1016/0893-6080\(89\)90020-8](http://dx.doi.org/10.1016/0893-6080(89)90020-8)

HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest mensuration**. 3. ed. New York: The Ronald Press Company, 1982. 410p.

JIANG, L.; BROOKS, J. R.; WANG, J. Compatible taper and volume equations for yellow-poplar in West Virginia. **Forest Ecology and Management**, Amsterdam, v. 213, p. 399-409, 2005. <http://dx.doi.org/10.1016/j.foreco.2005.04.006>

JOHANSSON, U.; BOSTRÖM, H.; LÖFSTRÖM, T.; LINUSSON, H. Regression conformal prediction with random forests. **Machine Learning**, v. 97, n. 1-2, p. 155-176, 2014. <http://dx.doi.org/10.1007/s10994-014-5453-0>

KAJISA, T.; MURAKAMI, T.; MIZOUE, N.; KITAHARA, F.; YOSHIDA, S. Estimation of stand volumes using the k-nearest neighbors method in Kyushu, Japan. **Journal of Forest Research**, v. 13, n. 4, p. 249-254, 2008. <http://dx.doi.org/10.1007/s10310-008-0077-5>

KOHLER, S. V. **Evolução do afilamento do tronco e do sortimento em plantios de *Pinus taeda* nos Estados do Paraná e Santa Catarina**. 84f. Dissertação (Mestrado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2013.

KOUKAL, T.; SUPPAN, F.; SCHNEIDER, W. The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. **Remote Sensing of Environment**, v. 110, n. 4, p. 431-437, 2007. <http://dx.doi.org/10.1016/j.rse.2006.08.016>

KOZAK, A.; MUNRO, D. D.; SMITH, J. G. H. Taper functions and their applications in forest inventory. **For Chron**, v. 45, n. 4, p. 278-283, 1969.

KOZAK, A. A variable-exponent taper equation. **Canadian Journal of Forest Research**, Ottawa, v. 18, n. 11, p.1363-1368, 1988. <http://dx.doi.org/10.1139/x88-213>

LARSON, P. R. Stern from development to forest tree. **Forest Science Monograph**, Washington, p.1-41, 1963.

LAWLER, J. J.; WHITE, D.; NEILSON, R. P.; BLAUSTEIN, A. R. Predicting climate-induced range shifts: model differences and model reliability. **Global Change Biology**, v. 12, n. 8, p. 1568-1584, 2006. <http://dx.doi.org/10.1111/j.1365-2486.2006.01191.x>

LEDUC, D. J.; MATNEY, T. G.; BELLI, K. L.; BALDWIN JR, V. C. Predicting diameter distributions of longleaf pine plantations: a comparison between artificial neural networks and other accepted methodologies. **Research Paper SRS – 25**, Southern Research Station, USA Department of Agriculture, Forest Service, 2001. Disponível em < http://www.srs.fs.usda.gov/pubs/rp/rp_srs025.pdf > Acesso em 11 de Agosto de 2015.

LEITE, H. G.; GARCIA, S. L. R. **Pesquisa e desenvolvimentos em inventário, mensurações e manejo florestal na CENIBRA**. Viçosa: Sociedade de Investigações Florestais, 2001. 49p. (Relatório técnico SIF).

LEITE, H. G.; SILVA, M. L. M.; BINOTI, D. H. B.; FARDIN, L.; TAKIZAWA, F. H. Estimation of inside-bark diameter and heartwood diameter for *Tectona grandis* Linn. trees using artificial neural networks. **European Journal of Forest Research**, v. 130, n. 2, p. 263-269, 2011. <http://dx.doi.org/10.1007/s10342-010-0427-7>

MACHADO, S. A.; URBANO, E.; CONCEIÇÃO, M. B.; FIGUEIREDO FILHO, A.; FIGUEIREDO, D. J. Comparação de modelos de afilamento do tronco para diferentes idades e regimes de desbastes em plantações de *Pinus oocarpa* Schiede. **Boletim de Pesquisa Florestal**, Colombo, n. 48, p.41-64, 2004.

MACHADO, S. A.; FIGUEIREDO FILHO, A. **Dendrometria**. 2. ed. Guarapuava: Unicentro, 2006. 316 p.

MALTAMO, M.; KANGAS, A. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. **Canadian Journal of Forest Research**, v. 28, n. 8, p. 1107-1115, 1998. <http://dx.doi.org/10.1139/x98-085>

MANSUY, N.; THIFFAULT, E.; PARÉ, D.; BERNIER, P.; GUINDON, L.; VILLEMAIRE, P.; POIRIER, V.; BEAUDOIN, A. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. **Geoderma**, v. 235-236, p. 59-73, 2014. <http://dx.doi.org/10.1016/j.geoderma.2014.06.032>

MAX, T. A.; BURKHART, H. E. Segmented polynomial regression applied to taper equations. **Forest Science**, v. 22, n. 3, p.283-289, 1976.

MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v.5, n. 4, p.115-133, 1943. <http://dx.doi.org/10.1007/BF02478259>

MCINERNEY, D. O.; NIEUWENHUIS, M. A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory. **International Journal of Remote Sensing**, v. 30, n. 19, p. 4937-4955, 2009. <http://dx.doi.org/10.1080/01431160903022936>

MCROBERTS, R. E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. **Forest Ecology and Management**, v. 272, p. 3-12, 2012. <http://dx.doi.org/10.1016/j.foreco.2011.06.039>

MCROBERTS, R. E.; NELSON, M.D.; WENDT, D.G. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. **Remote Sensing of Environment**, v. 82, n. 2-3, p. 457-468, 2002. [http://dx.doi.org/10.1016/S0034-4257\(02\)00064-0](http://dx.doi.org/10.1016/S0034-4257(02)00064-0)

MCROBERTS, R. E.; TOMPPO, E. O.; FINLEY, A. O.; HEIKKINEN, J. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. **Remote Sensing of Environment**, v. 111, n. 4, p. 466-480, 2007. <http://dx.doi.org/10.1016/j.rse.2007.04.002>

MCROBERTS, R. E.; NÆSSET, E.; GOBAKKEN, T. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. **Remote Sensing of Environment**, v. 163, p. 13-22, 2015. <http://dx.doi.org/10.1016/j.rse.2015.02.026>

MIGUEL, E. P. **Avaliação biométrica e prognose da produção de *Eucalyptus urophylla* (S.T. Blake) na região norte do estado de Goiás**. 161f. Dissertação (Mestrado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2009.

MITCHELL, T. M. **Machine Learning**. New York: McGraw Hill, 1997. 414p.

MOCHIUTTI, S. **Produtividade e sustentabilidade de plantações de acácia-negra (*Acacia mearnsii* De Wild.) no Rio Grande do Sul**. 270 f. Tese (Doutorado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2007.

MOGNON, F.; CORTE, A. P. D.; SANQUETTA, C. R.; BARRETO, T. G.; WOJCIECHOWSKI, J. Estimativas de biomassa para plantas de bambu do gênero *Guadua*. **Revista Ceres**, Viçosa, v. 61, n. 6, 2014. <http://dx.doi.org/10.1590/0034-737X201461060003>

MORENO, J. A. Clima do Rio Grande do Sul. **Boletim Geográfico do Rio Grande do Sul**, Porto Alegre, v. 11, p. 49-83, 1961.

MORENO-FERNÁNDEZ, D.; CAÑELLAS, I.; BARBEITO, I.; SÁNCHEZ-GONZÁLEZ, M.; LEDO, A. Alternative approaches to assessing the natural regeneration of Scots

pine in a Mediterranean forest. **Annals of Forest Science**, v. 72, n. 5, p. 569-583, 2015. <http://dx.doi.org/10.1007/s13595-015-0479-4>

MULLER, I. **Forma de tronco e sortimentos de madeira para *Eucalyptus grandis* Hill ex Maiden. manejado em alto fuste, na região sudeste do estado do Rio Grande do Sul.** 142 f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal de Santa Maria, 2004.

NOTHDURFT, A.; SABOROWSKI, J.; BREIDENBACH, J. Spatial prediction of forest stand variables. **European Journal of Forest Research**, v. 128, n. 3, p. 241-251, 2009. <http://dx.doi.org/10.1007/s10342-009-0260-z>

OHMANN, J. L.; GREGORY, M. J.; ROBERTS, H. M.; COHEN, W. B.; KENNEDY, R. E.; YANG, Z. Mapping change of older forest with nearest-neighbor imputation and Landsat time-series. **Forest Ecology and Management**, v. 272, p. 13-25, 2012. <http://dx.doi.org/10.1016/j.foreco.2011.09.021>

OHMANN, J. L.; GREGORY, M. J.; ROBERTS, H. M. Scale considerations for integrating forest inventory plot data and satellite image data for regional forest mapping. **Remote Sensing of Environment**, v. 151, p. 3-15, 2014. <http://dx.doi.org/10.1016/j.rse.2013.08.048>

ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; BROOKS, J. R.; WIAJANT JR, H. V. Estimating tree bole volume using artificial neural network models for four species in Turkey. **Journal of Environmental Management**, v. 91, n. 3, p. 742-753, 2010. <http://dx.doi.org/10.1016/j.jenvman.2009.10.002>

ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; CRECENTE-CAMPO, F.; ELER, U. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. **Forest Ecology and Management**, Amsterdam, v. 306, p. 52-60, 2013. <http://dx.doi.org/10.1016/j.foreco.2013.06.009>

ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; BROOKS, J. R. The use of tree crown variables in over-bark diameter and volume prediction models. **IForest**, v. 7, n. 3, p. 132-139, 2014. <http://dx.doi.org/10.3832/ifor0878-007>

PARRESOL, B. R.; HOTVEDT, J. E.; CAO, Q. V. A volume and taper prediction system for bald cypress. **Canadian Journal of Forest Research**, Ottawa, v. 17, p. 250-259, 1987. <http://dx.doi.org/10.1139/x87-042>

PELLI, E. **Comparação dos algoritmos máquina de aprendizagem extrema e retropropagação do erro para estimação de altura e volume de árvores.** 61 f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal dos Vales do Jequitinhonha e Mucuri, 2013.

PEREIRA, R. D. A. **Modelagem neuro-fuzzy na estimativa do perfil de *Eucalyptus* sp.** Monografia (Bacharel em Ciência da Computação) – Universidade Federal de Lavras, 2008.

PIERCE, A. D.; FARRIS, C. A.; TAYLOR, A. H. Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. **Forest Ecology and Management**, v. 279, p. 77-89, 2012. <http://dx.doi.org/10.1016/j.foreco.2012.05.010>

PRASAD, A. M.; IVERSON, L. R.; LIAW, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. **Ecosystems**, v. 9, n. 2, p. 181-199, 2006. <http://dx.doi.org/10.1007/s10021-005-0054-1>

PRODAN, M.; PETERS, R.; COX, F.; REAL, P. **Mensura forestal**. San Jose: IICA, 1997. 561p.

QUEIROZ, D.; MACHADO, S. A.; FIGUEIREDO FILHO, A.; ARCE, J. E.; KOEHLER, H. S. Identidade de modelos em funções de afilamento para *Mimosa scabrella* Bentham em povoamentos nativos da região metropolitana de Curitiba/PR. **Floresta**, Curitiba, v. 38, n. 2, p. 339-349, 2008. <http://dx.doi.org/10.5380/rf.v38i2.11629>

QUINLAN, R. J. Learning with Continuous Classes. In: 5th Australian Joint Conference on Artificial Intelligence, Singapore, 343-348, 1992.

REFENES, A. N.; ZAPRANIS, A.; FRANCIS, G. Stock performance modeling using neural networks: a comparative study with regression models. **Neural Networks**, v. 7, n. 2, p. 375-388, 1994. [http://dx.doi.org/10.1016/0893-6080\(94\)90030-2](http://dx.doi.org/10.1016/0893-6080(94)90030-2)

REZENDE, S. O. **Sistemas inteligentes**: fundamentos e aplicações. Barueri: Manole, 2003.

ROJO, A.; PERALES, X.; SÁNCHEZ-RODRÍGUEZ, F.; ÁLVAREZ-GONZÁLEZ, J. G.; VON GADOW, K. Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain). **European Journal of Forest Research**, Monique, v. 124, n. 3, p. 177-186, 2005. <http://dx.doi.org/10.1007/s10342-005-0066-6>

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagation of errors. **Nature**, London, v. 323, p.533-536, 1986b.

RUMELHART, D. E.; MCCLELLAND, J. L. A general framework for parallel distributed processing. In: Parallel distributed processing: explorations. In: Learning internal representations by error propagation. RUMELHART, D. E.; MCCLELLAND, J. L. (eds.). [S. l.]: MIT Press, 1986. p. 318-362.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence**: a modern approach. 3. ed. Upper Saddle River, N.J.: Prentice-Hall, 2010.

SANTOS, J. S. **Estratificação de florestas de eucalipto com base na forma do fuste das árvores**. 72 f. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2014.

SANQUETTA, C. R.; WOJCIECHOWSKI, J.; CORTE, A. P. D.; RODRIGUES, A. L.; MAAS, G. C. B. On the use of data mining for estimating carbon storage in the trees.

Carbon Balance and Management, v. 8, n. 6, 2013. <http://dx.doi.org/10.1186/1750-0680-8-6>

SANQUETTA, C. R.; WOJCIECHOWSKI, J.; CORTE, A. P. D.; BEHLING, A.; PÉLLICO NETTO, S.; RODRIGUES, A. L.; SANQUETTA, M. N. I. Comparison of data mining and allometric model in estimation of tree biomass. **BMC Bioinformatics**, 16:247, 2015. <http://www.biomedcentral.com/1471-2105/16/247>

SANQUETTA, C. R.; SANQUETTA, M. N. I.; CORTE, A. P. D.; PÉLLICO NETTO, S.; MOGNON, F.; WOJCIECHOWSKI, J.; RODRIGUES, A. L. Modeling the apparent volume of bamboo culms from Brazilian plantation. **African Journal of Agricultural Research**, v. 10, n. 42, p. 3977-3986, 2015b.

SAS Institute Inc. SAS/STAT User's Guide, Version 9.2. SAS Institute: Cary Inc. 2002.

SCHIKOWSKI, A. B.; CORTE, A. P. D.; SANQUETTA, C. R. Estudo da forma do fuste utilizando redes neurais artificiais e funções de afilamento. **Pesquisa Florestal Brasileira**, Colombo, v. 35, n. 82, p. 119-127, 2015. <http://dx.doi.org/10.4336/2015.pfb.35.82.867>

SCHRÖDER, T.; PEREIRA, L. D.; HOFIÇO, N. S. A. Comparação de Métodos de Estimativa de Volume Total para *Eucalyptus grandis* W. Hill ex Maiden. **Floresta e Ambiente**, Rio de Janeiro, v. 20, n.4, p. 480-486, 2013. <http://dx.doi.org/10.4322/floram.2013.041>

SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, v.47, n.9, p.719-734, 1933.

SCORNET, E. On the asymptotics of random forests. **Journal of Multivariate Analysis**, 2015. No prelo. <http://dx.doi.org/10.1016/j.jmva.2015.06.009>

SEYEDHOSSEINI, M.;TASDIZEN, T. Disjunctive normal random forests. **Pattern Recognition**, v. 48, n. 3, p. 976-983, 2015. <http://dx.doi.org/10.1016/j.patcog.2014.08.023>

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANI, J. M.; LEITE, H. G. Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, Viçosa, v. 33, n. 6, p. 1133-1139, 2009. <http://dx.doi.org/10.1590/S0100-67622009000600015>

SILVA, F.; CORTE, A. P. D.; SANQUETTA, C. R. Equações de afilamento para descrever o volume total do fuste de *Pinus caribaea* var. *hondurensis* na região do Triângulo Mineiro. **Scientia Forestalis**, Piracicaba, v. 39, n. 91, p. 367-376, 2011.

SIRONEN, S.; KANGAS, A.; MALTAMO, M.; KANGAS, J. Estimating individual tree growth with nonparametric methods. **Canadian Journal of Forest Research**, v. 33, n. 3, p. 444-449, 2003. <http://dx.doi.org/10.1139/x02-162>

SOARES, F. A. A. M. N.; FLÔRES, E. L.; CABACINHA, C. D.; CARRIJO, G. A.; VEIGA, A. C. P. Recursive diameter prediction and volume calculation of Eucalyptus trees using multilayer perceptron networks. **Computers and Electronics in Agriculture**, v. 78, n. 1, p. 19-27, 2011.

<http://dx.doi.org/10.1016/j.compag.2011.05.008>

SOARES, F. A. A. M. N.; FLÔRES, E. L.; CABACINHA, C. D.; CARRIJO, G. A.; VEIGA, A. C. P. Recursive diameter prediction for calculating merchantable volume of Eucalyptus clones without previous knowledge of total tree height using artificial neural networks. **Applied Soft Computing**, v.12, p. 2030-2039, 2012.

<http://dx.doi.org/10.1007/s00521-012-0823-7>

SOARES, F. A. A. M. N.; FLÔRES, E. L.; CABACINHA, C. D.; CARRIJO, G. A.; VEIGA, A. C. P. Recursive diameter prediction for calculating merchantable volume of eucalyptus clones using Multilayer Perceptron. **Neural Computing and Applications**, v. 22, n. 7-8, p. 1407-1418, 2013. <http://dx.doi.org/10.1007/s00521-012-0823-7>

SOUZA, C. A. M. **Modelos de afilamento para *Pinus taeda* L. ajustados segundo a forma do tronco e métodos de estratificação**. 120 f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal de Santa Maria, Santa Maria, 2009.

SOUZA, R. R. **Estudo da forma do fuste de árvores de eucalipto em diferentes espaçamentos**. 86 f. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2013.

STROBL, C.; BOULESTEIX, A. L.; KNEIB, T.; AUGUSTIN, T.; ZEILEIS, A. Conditional variable importance for random forests. **BMC Bioinformatics**, v. 9, p.307, 2008. <http://dx.doi.org/10.1186/1471-2105-9-307>

STROBL, C.; MALLEY, J.; TUTZ, G. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. **Psychol Methods**, v. 14, p.323-348, 2009. <http://dx.doi.org/10.1037%2Fa0016973>

TATSUMI, K.; YAMASHIKI, Y.; TORRES, M. A. C.; TAIPE, C. L. R. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. **Computers and Electronics in Agriculture**, v. 115, p. 171-179, 2015. <http://dx.doi.org/10.1016/j.compag.2015.05.001>

TÉO, S. L.; MARCON, A.; EHLERS, T.; BIANCHI, J. C.; PELOSO, A.; NAVA, P. R.; COSTA, R. H. Modelos de afilamento para *Pinus elliottii* em diferentes idades, na região de Caçador, SC. **Floresta**, Curitiba, v. 43, n. 3, p. 439 - 452, 2013.

TRAMONTANA, G.; ICHII, K.; CAMPS-VALLS, G.; TOMELLERI, E.; PAPALE, D. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. **Remote Sensing of Environment**, v. 168, p. 360-373, 2015. <http://dx.doi.org/10.1016/j.rse.2015.07.015>

TURING, A. M. Computing machinery and intelligence. **Mind**. v. 59, p. 433-460, 1950.

WANG, Y.; WITTEN, I. H. Induction of model trees for predicting continuous classes. **WorkingPaper 96/23**. 1996. Disponível em:<<http://www.cs.waikato.ac.nz/pubs/wp/1996/uow-cs-wp-1996-23.pdf>>. Acesso em 22/06/2015.

WERE, K.; BUI, D. T.; DICK, O. B.; SINGH, B. R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, v. 52, p. 394-403, 2015. <http://dx.doi.org/10.1016/j.ecolind.2014.12.028>

WILSON, B. T.; LISTER, A. J.; RIEMANN, R. I. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. **Forest Ecology and Management**, v. 271, p. 182-198, 2012. <http://dx.doi.org/10.1016/j.foreco.2012.02.002>

WITTEN, I. H.; EIBE, F.; HALL, M. A. **Data mining**: practical machine learning tools and techniques.3.ed. Burlington: Elsevier, 2011. 665p.

WOJCIECHOWSKI, J. **JCarbon - software na web com data mining para estimativas de volume, biomassa e carbono em florestas**. 174 f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2015.

YOSHITANI JUNIOR, M.; NAKAJIMA, N. Y.; ARCE, J. E.; MACHADO, S. A.; DRUSZCZ, J. P.; HOSOKAWA, R. T.; MELLO, A. A. Funções de afilamento para plantios desbastados de *Pinus taeda*. **Floresta**, Curitiba, v. 42, n. 1, p. 169-176, 2012.